

Kognícia a umelý život 2017

zostavili

Igor Farkaš
Martin Takáč
Ján Rybár
Peter Gergel'

Univerzita Komenského v Bratislavе - 2017



Akademie věd
České republiky
Strategie AV21
Společný výzkum ve veřejném zájmu



nuabi
internet services

Hotel Flóra
Trenčianske Teplice
31. 5.-2. 6. 2017

Kognícia a umelý život 2017
(recenzovaný zborník)

Vznik tejto publikácie bol podporený Kultúrnou a edukačnou agentúrou Ministerstva školstva, vedy a športu Slovenskej republiky (KEGA), v rámci projektu č. 017UK-4/2016.

Recenzenti príspevkov:

prof. RNDr. Lubica Beňušková, PhD.
RNDr. Barbora Cimrová, PhD.
RNDr. Zuzana Černeková, PhD.
prof. Ing. Igor Farkaš, Dr.
MUDr. Mgr. Tomáš Hromádka, PhD.
Ing. Matěj Hoffmann, PhD.
PhDr. Juraj Hvorecký, PhD.
Mgr. Dalibor Jurášek
PhDr. Dezider Kamhal, PhD.

Mgr. Lenka Kostovičová, PhD.
RNDr. Andrej Lúčny, PhD.
Mgr. Martin Marko, PhD.
doc. RNDr. Mária Markošová, PhD.
RNDr. Marek Nagy, PhD.
prof. RNDr. Jiří Pospichal, PhD.
doc. PhDr. Ján Rybár, PhD.
RNDr. Martin Takáč, PhD.

Vydavateľ:
Univerzita Komenského v Bratislave
Šafárikovo nám. 6
81000 Bratislava

Zostavili ©
prof. Ing. Igor Farkaš, Dr., 2017
RNDr. Martin Takáč, PhD., 2017
doc. PhDr. Ján Rybár, PhD., 2017
Mgr. Peter Gergel, 2017

Autorské práva © autori príspevkov, 2017

Dizajn obálky © Róbert Kelemen a Kristína Malinovská, 2017

Tlač: Univerzita Komenského v Bratislave

Vydané v Bratislave v máji 2017

ISBN 978-80-223-4346-6

Predslov

Milé kolegyne a kolegovia,

máte pred sebou už 17. vydanie zborníka konferencie Kognícia a umelý život. Publikácia vznikla hlavne zásluhou vás, odborníkov z Česka a Slovenska, ktorí ste reflektovali na toto podujatie a rozhodli sa podporiť ho svojou účasťou. Tomuto sa veľmi tešíme, pretože počet záujemcov sympaticky narastal a mierne prekročil okrúhlu päťdesiatku.

Po úspešnej premiére pred dvoma rokmi sa KUŽ po druhýkrát koná v Trenčianskych Tepliciach v hoteli Flóra. Sme presvedčení, že sa nám opäť podarí vytvoriť tvorivú atmosféru, ktorá podnieti zaujímavé interdisciplinárne diskusie nielen počas sekcií, ale aj mimo nich. Kognitívna veda je svojou podstatou programovo zameraná na diskusie naprieč jednotlivými relevantnými odbormi, ponúka možnosť pozrieť sa aj za hranice vlastnej oblasti, nachádzať rôzne súvislosti a analógie. Dúfame, že v publikovaných príspevkoch v tomto zborníku nájdete inšpiračné zdroje pre vaše ďalšie skúmanie v oblastiach vášho záujmu. Rozhodli sme sa pokračovať v eko-prístupe, veriac, že zborník na klúči bude vnímaný pozitívne.

Do programu tohoročnej konferencie sme zaradili tri zaujímavé pozvané odborné prednášky. Prof. Ľubica Beňušková (Centrum pre kognitívnu vedu FMFI UK) nás zasväti do tajov funkčných sietí mozgu a ich zmien v dôsledku starnutia človeka. Doc. Ivan M. Havel sa bude venovať výskumu z perspektívy prvej osoby a oboznámi nás so svojimi poznatkami týkajúcimi sa introspekcie. MUDr. Igor Riečanský (Ústav normálnej a patologickej fyziológie SAV v Bratislave a Katedra psychológie Viedenskej univerzity) sa bude venovať sociálnej kognícii a úlohe zdieľaných mozgových aktivácií.

Program KUŽ 2017 pokrýva spektrum príspevkov od robotiky, cez neurovedu, psychológiu, jazykovedu, strojové učenie, až po estetiku. Ide o práce teoretického aj experimentálneho charakteru. Každý príspevok prešiel recenzným konaním, ktoré prebehlo veľmi úspešne vďaka poctivej práci všetkých oslovených recenzentov. Za to im srdečne ďakujeme.

Vďaka vyššiemu počtu prijatých príspevkov sme sa rozhodli obohatiť program posterovou sekciou, ktorá umožňuje dlhšie interakcie autorov so záujemcami.

Veľké podčiarkovanie patrí členom Centra pre kognitívnu vedu na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, ktorí vyvinuli potrebné úsilie pri príprave konferencie. Osobitne oceňujeme prácu našich kolegýň Kristíny Malinovskej a Barbory Cimrovej a študentky Miroslavy Galasovej. Spolupráca s nimi bola potešením.

Tohoročné vložné pre študentov sme výrazne znížili vďaka podpore z programu Stratégia 21 Akadémie vied ČR. Okrem toho, na konferencii KUŽ 2017 sa mohli aktívne zúčastniť aj traja motivovaní študenti, ktorých vložné pokryla Slovenská spoločnosť pre kognitívnu vedu, a jeden študent bol podporený firmou Nuabi, s.r.o., za čo ďakujeme pánovi Davidovi Navaroví.

Vydanie konferenčného zborníka bolo podporené Kultúrnou a edukačnou agentúrou Ministerstva školstva, vedy a športu Slovenskej republiky (KEGA) v rámci projektu č. 017UK-4/2016.

Bratislava, máj 2017

Igor Farkaš, Martin Takáč, Ján Rybár, Peter Gergel'

Obsah

Pozvaní rečníci	7
Rečové prispôsobovanie sa medzi človekom a automatickým systémom Štefan Beňuš	8
Effect of aging on functional brain networks: a graph-theoretical analysis <i>Lubica Beňušková, Paul McCarthy, Elizabeth A. Franz</i>	14
Etické a legislatívne problémy využívania dronov Ivana Budinská	24
Racionálni učitelia a intuitívni manažéri: Interakcia veku a kognitívnej reflexie pri rozhodovaní Vladimíra Čavojová, Marek Jurkovič	30
Využitie Gärdenforsovej geometrie významu pre uchopenie sémantiky pojmu krásy Andrej Démuth	37
Redukcia kognitívnych odchýlok u finančných profesionálov prostredníctvom kontrafaktov Katarína Dudeková, Lenka Kostovičová, Matúš Konečný	46
Ja na rozhraní vedy Silvia Gálíková	52
Improving the performance of impulse neuro-glial network Peter Gergel'	60
Vnímání pohybu: Reflektovaný a nereflektovaný pohled v souvislosti s architekturou Ivan M. Havel, Monika Mitášová	64
Vývoj teórie myслe u detí a adolescentov Martin Jáni	71
Computational models of homeostatic synaptic and dendritic plasticity in the hippocampus Peter Jedlička	76
Primitiva virtuální interakce: základní stavební kameny pro vývoj a výzkum interaktivních virtuálních prostředí Vojtěch Juřík, Pavel Ugwitz, Čeněk Šašinka	81

Teória rečových aktov a jej implikácie <i>Dezider Kamhal</i>	87
Vplyv iracionálneho presvedčenia na úspešnosť v riešení úloh u finančných profesionálov <i>Lenka Kostovičová, Katarína Dudeková, Matúš Konečný</i>	93
Graphical Granger Causality for Detection Temporal Anomalies in EEG Data <i>Anton Kováč, Kateřina Hlaváčková-Schindler</i>	99
Robotic reaching and grasping using a TRPO-based reinforcement learning <i>Peter Kovács, Igor Farkaš</i>	104
Metafora a literárna recepcia <i>Jana Kuzmíková</i>	110
A Model of the Reference Frame of the Ventriloquism Aftereffect using a priori bias <i>Peter Lokša, Norbert Kopčo</i>	116
Robot pred zrkadlom <i>Andrej Lúčny</i>	123
Replicator and ART2 Neural Networks in Text Outliers Identifications <i>Asmaa Salem, Gabriela Andrejková, Pavol Kozák</i>	130
Auditory spatial discrimination with visual vs. auditory attentional cueing <i>René Šebeňa, Bernadeta Hrebeňárová, Norbert Kopčo</i>	136
Exploring the Neural Correlates of Different Interaction Types: a Hyperscanning Investigation Using the Pattern Game <i>Beáta Šplíšková, Daniel J. Shaw</i>	141
Bariéry (nielen) psychologickej vedy: Kognitívne odchýlky vo vedeckom myslení <i>Jakub Šrol</i>	149
Where is my forearm? Clustering of body parts from simultaneous tactile and linguistic input using sequential mapping <i>Karla Štěpánová, Matěj Hoffmann, Zdeněk Straka, Frederico B. Klein, Angelo Cangelosi, Michal Vavrečka</i>	155
Modelování života a organismus uměleckého díla <i>Aleš Svoboda</i>	163
Role třídění z hlediska stability v evoluci <i>Jan Toman, Jaroslav Flegr</i>	169
Effect of stimulus distribution on the buildup of contextual plasticity in sound localization <i>Beáta Tomoriová, Gabriela Andrejková, Norbert Kopčo</i>	178

Visual cognitive styles in virtual environments: Constructing and evaluating adequate task <i>Pavel Ugwitz, Vojtěch Juřík, Čeněk Šašinka</i>	184
Nový pohled na výpočty a umělá inteligence <i>Jiří Wiedermann</i>	190
Register autorov	194

Pozvaní rečníci

prof. Lubica Beňušková: Effect of aging on functional brain networks: a graph-theoretical analysis



Graph theory is a novel approach to the analysis of functional connectivity in the brain. We have performed graph theoretical analysis of an fMRI data set from a study conducted by Buckner et. al. (2000), in an attempt to find differences between healthy participants, both young and aged, and aged participants with mild Alzheimer's Disease. We apply our techniques to the data, and also give a preliminary statistical comparison across these groups. Our results show that there are statistically significant differences between the groups in several global network characteristics, including density, number of nodes and maximum degree. We hypothesise that these differences reflect a general change in how the aging brain copes with everyday tasks, meaning an increased effort required by aged brains to perform, which is further accentuated by the onset of Alzheimer's disease.

doc. Ivan M. Havel: Zápisky introspektora



Přednáška se zabývá metodou zkoumání mysli a vědomí, která je na rozhraní kognitivní vědy a filosofické fenomenologie. Navrhovaná metoda, nazvaná "introspekcí plus," se zaměřuje na subjektivní prožitky velmi krátkého trvání, jako jsou například mikrosy, vidiny či mžikové vizuální vjemů. Mentální obrazy takových prožitků setrvávají ve vědomí i krátce po skončení, takže mohou být retrospekčně analyzovány z perspektivy "první osoby" prožívajícího subjektu.

MUDr. Igor Riečanský: Zdieľané mozgové aktivácie a sociálna kognícia



Príspevok sa zaobrá fenoménom zdieľaných aktivácií mozgu, nazývaných aj rezonantné aktivácie. Jedná sa o stavy prieniku v aktivácii mozgových procesov medzi jedincom nachádzajúcim sa v určitom psychofyziologickom stave a jedincom, ktorý ho pozoruje, t.j. vníma iného jedinca nachádzajúceho sa v danom stave. Existenciu zdieľaných aktivácií potvrzuje množstvo experimentálnych neurozobrazovacích aj elektrofyziológických štúdií a novšie aj štúdie s použitím cielenej intervencie, napr. farmakologickej manipulácie psychických a mozgových procesov. Výskumy u zdravých jedincov aj osôb s duševnými poruchami svedčia o tom, že zdieľané aktivácie by mohli hrať významnú úlohu v sociálnej kognícii.

Rečové prispôsobovanie sa medzi človekom a automatickým systémom

Štefan Beňuš

Univerzita Konštantína Filozofa v Nitre, Štefánikova 67 94974 Nitra
Ústav Informatiky SAV, Dúbravská cesta 9, 845 07 Bratislava
sbenus@ukf.sk

Abstrakt

Základnou otázkou tohto výskumu je to, akým hlasom by mal rozprávať robot alebo komunikačný systém, aby jeho príkazy a rady boli dôveryhodné. Pri medziľudskej komunikácii dochádza k prispôsobovaniu sa medzi rečníkmi prirodzené a je spojené s kladným hodnotením spolurečníka, úspešnosti komunikácie, a ďalšími pozitívami. Vychádzajúc z tohto poznania sme vyvinuli komunikačnú hru, v ktorej človek hrá kartovú hru a počas nej dostáva rady od avatarov. Skúmame, či hráč bude viac dôverovať radám od toho avatara, ktorého syntetický hlas sa prozodicky prispôsobuje hlasu hráča. Doterajšie výsledky naznačujú, že rečové prispôsobovanie automatického systému človeku môže ovplyvňovať dôveryhodnosť systému a aj rozhodovanie človeka, ktorý so systémom komunikuje.

1 Úvod

Komunikácia medzi strojom a človekom v súčasnosti už bežne prebieha v rečovej modalite, čo je možné doložiť úspešnosťou a oblúbenosťou komunikačných asistentov ako Siri od Apple, Cortana od Microsoftu, Alexy od Amazonu, alebo Google Assistant. Tento boom bol umožnený veľkým progresom v počítačovom spracovaní hlasového signálu pri rozpoznávaní ľudského hlasu a generovaní syntetického hlasu. Charakteristiky dialógov medzi človekom a strojom však ešte ani zdáleka nedosahujú komplexnosť medziľudskej komunikácie. Komplexnosť vztáhov, ktoré sa dialógmi vytvárajú a upevňujú, je teda omnoho väčšia medzi ľuďmi ako pri komunikácii človek-stroj.

Jednou z oblastí, kde sa tento rozdiel prejavuje je komunikatívne prispôsobovanie sa, ktoré sa v literatúre rôznych vedeckých oblastí označuje rôzne, ako napríklad entrainment, accommodation, adaptation, alignment, synchrony, a inak, pozri napr. prehľadové štúdie Beňuš (2014) alebo Brannigan a kol. (2011). V rečovej medziľudskej interakcii bolo takéto prispôsobovanie preukázané v mnohých lingvistických ako aj paralingvistických rovinách dialógu, napríklad vo fonetickej realizácii jednotlivých slov (Pardo 2006), výbere lexikálnych jednotiek a syntaktických štruktúr (Brennan 2000, Reitter a kol. 2007, Brannigan a kol. 2000), časových charakteristikách preberania slova (ten

Bosch a kol. 2005, Beňuš a kol. 2011), alebo akusticko-prozodických vlastnostiach ako výška a intenzita hlasu alebo tempo reči (napr. Levitan a Hirschberg 2011).

Komunikatívne prispôsobovanie medzi spolurečníkmi je prepojené na sociálnu rovinu medziľudskej komunikácie, keďže existuje silná korelácia medzi mierou prispôsobovania a pozitívneho vnímania spolurečníka a aj celkovej interakcie. Táto myšlienka je formalizovaná v Communication Accommodation Theory (CAT, Giles a kol. 1991), ktorá hovorí, že miera prispôsobovania je jednou z možností ako dosiahnuť (alebo ovplyvniť) mieru vzájomnej blízkosti medzi spolurečníkmi. V princípe, väčšia miera prispôsobovania viedie k pozitívnejšiemu výsledku komunikácie a rečové prispôsobovanie tak odzrkadľuje snahu o sociálnu integráciu.

V poslednej dobe v snahe využiť tieto pozitívne črty prispôsobovania a vylepšiť automatické dialógové systémy sa formy komunikatívneho prispôsobovania začali skúmať aj pri interakcii človek-stroj (napr. Bell a kol. 2003, Brannigan a kol. 2010). Implementácia prispôsobovania sa automatického systému k svojmu užívateľovi má niekoľko výhod. Na jednej strane je dobré ak sa automatický systém snaží implementovať črty medziľudskej dialógu aby sa tak zvyšovala prirodzenosť dialógových systémov. Pri aplikáciách zameraných na poskytovanie služieb alebo plnenie úloh môže takéto prispôsobovanie pozitívne ovplyvňovať vnímanie užívateľa smerom k takému systému ako aj vnímanie úspešnosti celej interakcie podobne, ako to bolo pozorované v medziľudskej interakciach. Pri aplikáciách, kde automatický systém slúži ako konverzačný spoločník, asistent pri doučovaní, alebo spoločník seniorov s medicínskymi a spoločenskými úlohami, môže takéto prispôsobovanie pozitívne ovplyvniť spriaznenosť užívateľa s takýmto systémom, čo následne môže viesť k väčšej efektivite pri plnení úloh. V nedávnej štúdii napríklad Lubold a kol. (2016) zistili, že akusticko-prozodické prispôsobovanie sa robotického pomocníka pri doučovaní signifikantne zvyšuje "social presence", teda vnímanie interakcie s takýmto robotickým spoločníkom ako prirodzené, príjemné, a motivujúce.

Z kognitívneho hľadiska je samozrejme výskum vzájomného rečového prispôsobovania sa dôležitý pri formulácii modelov kognitívneho systému reči, ktoré donedávna boli založené na poznatkoch hlavne z

monologického jazykového prejavu, ale modely kognitívnych jazykových schopností založené na podstate dialógu sa v poslednej dobe dostávajú do popredia (napr. Pickering a Garrod 2013). Popri výskume kognitívneho rečového systému vstupuje do popredia aj výskum rozhodovacích procesov človeka pri jeho komunikácii so strojom a ako toto rozhodovanie môže byť ovplyvnené.

Je potrebné však spomenúť, že prispôsobovanie a jeho potenciálne pozitívne črty sa nemusia v interakcii medzi človekom a strojom objavovať automaticky. Strupka a kol. (2016) napríklad zistili, že pri komunikácii s robotmi, ktorí hovorili mužskými a ženskými hlasmi, užívatelia prispôsobili intenzitu a intonáciu hlasu pri rozdielnom pohlaví robotického hlasu, ale efekt bol opačný: hlasové charakteristiky užívateľov sa nepribližovali robotovým, ale sa od nich vzďaľovali.

Smer výskumu popísaný vyššie by sme radi posunuli preskúmaním možnosti, že akusticko-prozodické prispôsobovania automatických syntetických hlasov ovplyvňuje dôveryhodnosť užívateľa voči týmto hlasom a môže na základe zmeny dôveryhodnosti aj ovplyvniť ich rozhodovanie.

2 Metódy

2.1 Prispôsobovanie

Spôsobov, ako je možné rečové prispôsobovanie v dialógoch uchopiť, je niekoľko. Levitan a Hirschberg (2011) popísali tri základné: podobnosť, konvergencia, a synchrónnosť. Pri podobnosti ide o porovnanie priemernej hodnoty určitej charakteristiky, napríklad tempa reči, počas celého dialógu medzi rečníkmi. Môže sa to využiť napríklad na porovnanie takto formulovanej miery podobnosti medzi rečníkom a jeho/jej spolurečníkom na jednej strane a na druhej strane medzi rečníkom a "imaginárnym spolurečníkom", teda napríklad priemernou hodnotou u ostatných hovoriacich v databáze, ktorí však s rečníkom nekomunikovali. Ak je rozdiel rečník-spolurečník menší, ako rečník-imaginárny spolurečník, došlo k rečovému prispôsobaniu sa medzi rečníkom a spolurečníkom.

Pri konvergencii ide o mieru približovania sa hodnôt určitej rečovej charakteristiky medzi spolurečníkmi počas celého dialógu. Je to teda čiastočne dynamická veličina (aj keď vyjadrená statickou jednou hodnotou z celého rozhovoru) miery časovej zmeny podobnosti.

Poslednou formou uchopenie prispôsobovania je synchrónnosť. Toto je najviac dynamický koncept, ktorý vyjadruje relatívnu koordináciu spolurečníkov pri každej zmene hovoriaceho, teda pri každom prevzatí slova (turn exchange). Nejde tu teda o konvergenciu, keďže "vzdialenosť" medzi spolurečníkmi sa nemusí zmenšovať, ale o prispôsobovanie sa smeru zmeny. Napríklad, ak Rečník 1 hovoril rýchlejšie ako Rečník 2

a v určitem momente Rečník 1 spomalí, Rečník 2 sa bude správať synchrónne, ak aj on spomalí a nepribliží sa tak Rečníkovi 1, ale jeho správanie bude konvergentné, ak trochu zrýchli, aby sa tak "vzdialenosť" medzi rečníkmi zmenšila.

Ako forma prispôsobovania v našej komunikatívnej hre opísanej nižšie bola zvolená práve synchrónnosť. Hlavnými dôvodmi bolo to, že synchrónnosť sa hojne vyskytuje v medziľudskej komunikácii (napr. Edlund a kol. 2009, Levitan, a kol. 2015), a umožňuje aktívnejší a variabilnejší prístup k manipulácii syntetických hlasov počas hrania hry v porovnaní s konvergenciou alebo podobnosťou.

2.2 Komunikatívna hra Kvarteto s Pomocníkmi

2.2.1 Myšlienka

Hra Kvarteto je prispôsobenou verziou hry *GoFishwithHelpers*, ktorú navrhla vo svojej dizertačnej práci R. Levitan (2014). Hra testuje základnú otázku tohto výskumu, či hráč dôveruje viac tomu avatarovi, ktorého hlas sa mu prispôsobuje ako avatarovi, ktorého hlas ostáva statický, alebo sa aktívne neprispôsobuje tým, že sa pohybuje asynchronne voči hlasu hráča.

Použitá bola jednoduchá hra Kvarteto, kde si hráč pýta od svojho súpera kartu určitej hodnoty s cieľom zozbierať všetky štyri karty tejto hodnoty a vytvoriť tak Kvarteto. Úlohou hráča je nazbierať čo najviac bodov ktoré sa udeľujú pri získaní každej karty od súpera (10 bodov), pri získaní kvarteta (100 bodov), a 50 bodov sa stráca v situácii ak si hráč vyžiada kartu a súper žiadnu kartu tejto hodnoty nemá. Súperom hráča je počítač, ktorého karty hráč nevidí. Myšlienka využitia kartovej hry ovláданej hlasom na rôzne účely v slovenčine nie je nová (napr. Nagy 2006 využíva kvarteto na zlepšenie povedomia o čísliciach pre žiakov základných škôl). V tejto hre ale pri každom tahu hráč musí hlasovo požiadať o radu od jedného z dvoch avatarov, akú kartu by si mal od súpera vypýtať. Hráč vidí na obrazovke obrázky dvoch jasne rozdielnych ženských tvári (Monika a Tereza) a príklady otázok, ktoré môže použiť, napr. "Monika, ktorú kartu by si mi odporučila?". Rozpoznávač zistí z tejto otázky identitu avatara, ktorý bol o radu požiadany, a pri následnej rade sa ukáže obrázok požiadaneho avatara a syntetizátor prehrá danú radu, napr. "Ja by som odporučila šestku". Hlas jedného z avatarov sa počas celej hry aktívne dynamicky prispôsobuje hráčovi, t.j. každá jeho rada je podobná prozodickým charakteristikám otázky, ktorou si hráč radu vypýtal, a hlas druhého z avatarov sa hlasu hráča neprispôsobuje.

Dôležitým predpokladom testovania základnej výskumnej otázky je to, aby kvalita rady, ktorú avatar poskytne, neslúžila hráčovi na rozhodovanie, komu bude viac dôverovať. To znamená, že chceme predísť tomu, aby sa hráč rozhodoval podľa kvality rád. Toto je zabezpečené dvomi súčasťami hry. Prvou je

algoritmickej implementácie, ktorá zabezpečuje, že rozdiel výsledných bodových ziskov hráča plynúcich z rôznych avatarov je počas hry udržiavaný na minime, čo znamená, že obidva avatari dávajú v priemere celej hre rovnako dobré rady. Druhou je to, že hráč nevidí karty protihráča (počítača), ktorý pri každom ťahu dostane novú sadu kariet. Tým je možné, že pri každej rade existuje v princípe "lepšia" rada, ktorá by hráčovi priniesla viac bodov, ako aj "horšia" rada, ktorá by priniesla bodov menej. Tieto vlastnosti hry umožňujú, aby hráč, aj napriek tomu, že algoritmickej oba avatari poskytujú rady približne rovnakej kvality, mohol (mylne) predpokladať, že jeden z avatarov predsa len poskytuje lepšie rady.



Obr. 1: Grafické užívateľské rozhranie pre hru Kvarteto s pomocníkmi. Vpravo hore je tlačidlo na spustenie a zastavenie rozpoznávania (pri stlačení signalizované ikonou mikrofónu), pod ním obrázky avatarov a návrhy otázok, vľavo hore karty súpera (počítača) a vľavo dole karty hráča s aktuálnym skóre.

2.2.2 Implementácia

Naša implementácia je pozmenená oproti tej, ktorú použila Levitan (2014) vo viacerých ohľadoch, ktoré popíšeme nižšie. Využíva sa programovacie prostredie Python a http komunikácia so serverom zabezpečujúcim rozpoznávanie reči hráča a syntézu reči avatarov. Grafické užívateľské prostredie je ilustrované na Obr. 1 spolu s popisom jednotlivých súčasti.

Základom implementácie je extrakcia prozodických charakteristik reči s hlasu hráča a následná reprodukcia týchto charakteristik v hlate prispôsobujúceho sa avataru. V tomto príspevku prezentujeme výsledky dvoch experimentov: v prvom sa prispôsobovalo len tempo reči a v druhom priemerná intenzita, výška hlasu (f_0) a tempo reči.

Tempo reči je možné získať buď priamo z akustického signálu odhadom počtu slabík (napríklad použitím rozšíreného softvéru na fonetickú analýzu Praat (Boersma a Weenink 2017) a dostupných skriptov na takéto určenie tempa reči v Praat-e (Mertens 2014, de Jong a Wempe 2009), alebo

použitím výstupu z rozpoznávača, kde je počet slabík v rozpoznanom teste možné algoritmickej odvodiť podľa počtu slabičných elementov. Pri experimentálnom testovaní bola metóda použitia výstupu rozpoznávača viac spoľahlivá a bola teda implementovaná v hre Kvarteto. Priemerné hodnoty výšky hlasu a intenzity sa použili z výstupu Praat skriptu (de Jong a Wempe 2009), ktorý sme mierne prispôsobili našim potrebám.

Podstatným rozdielom medzi súčasnou implementáciou a tou, ktorú použila Levitan (2014) je spôsob prispôsobovania, teda reprodukcie prozodických vlastností reči hráča v syntetickej reči prispôsobujúceho sa avataram. Levitan zvolila reprodukciu absolútnych nameraných hodnôt z hlasu hráča v hlate avatara. V našich experimentoch sme zvolili relatívne prispôsobovanie synchronnosti popísanej v podkapitole 2.1. Pri každom ťahu si hráč vypýta radu, vypočítava tempo reči, výšku a intenzitu hlasu, tieto hodnoty sa porovnajú s hodnotami v otázke v predchádzajúcom ťahu, a v syntéze následnej rady prispôsobujúceho sa avatara sa posunú o tento rozdiel. Ak hráč zrýchli alebo hovoril hlasnejšie oproti svojmu predchádzajúcemu ťahu, prispôsobivý avatar tiež zrýchli alebo bude hovoriť hlasnejšie.

Pri implementácii hlasu avatara, ktorý sa neprispôsobuje sme testovali dva spôsoby. V prvom (*Adapt-Stable*) sa hlas avatara vôbec nehýbal a počas celej hry produkoval predvolené hodnoty daného syntetickejho hlasu. V druhom (*Adapt-Antiadapt*) sa hodnoty v syntéze neprispôsobivého hlasu posúvali v opačnom smere oproti zmene nameranej v hlate hráča: ak si hráč vypýtal radu vyšším hlasom ako v minulom ťahu o napr. 5 Hz, neprispôsobivý avatar poskytol radu hlasom o 5Hz zníženou f_0 .

Syntéza a rozpoznávanie bolo zabezpečené systémom vytvoreným na Ústave Informatiky SAV. Rozpoznávač využíva voľne dostupný dekódér reči Julius (<http://julius.osdn.jp/en-index.php>) obohatený o trifónový akustický model a jazykový model slovenčiny. Syntetizátor slovenčiny je založený na štatistickej parametrickej metóde voľne dostupného softvéru HTS (<http://hts.sp.nitech.ac.jp>) vytvorenéj z jedného ženského hlasu hovorennej slovenčiny na základe databázy čítaných textov.

V tejto implementácii je dialóg zjednodušený tak, že rozpoznávač len zistí identitu cieľového avatara (Monika-Tereza), a nápoveda sa vyberie náhodne z viacerých prednastavených viet. Hráč nemá možnosť sám si vybrať karty a radu avatara musí poslúchnuť. Jeho jedinou voľbou a stratégou je to, ktorého avatara sa opýta.

3 Experiment 1

3.1 Probanti

V experimente sa zúčastnilo 20 probantov. Všetci boli študenti nitrianskych univerzít, z toho 14 žien a 6 mužov.

3.2 Procedúra

Probanti sa zúčastnili experimentu po vyslovení informovaného súhlasu a boli finančne odmenení za svoj čas. Na začiatku vyplnili sociometrický dotazník a skrátený osobnostný test TIPI (Ten Item Personality Inventory, Gosling a kol. 2003). Potom sa oboznámili s pravidlami hry a inštrukciou, že pri hre budú komunikovať s dvomi avatami, ktorí väčšinou dávajú výhodné rady, ale niekedy môžu robiť aj chyby a ich úlohou v experimente je nahrať čo najviac bodov a tak zisťovať, ktorý z avatarov poskytuje lepšie rady.

Po úvodnom oboznámení sa s hrou a grafickým užívateľským rozhraním hráči pristúpili k samotnej hre s dvomi avatarmi (Monikou a Terezou) v celkovej dĺžke 45 ľahov. Na záver probanti vyplnili krátky dotazník, v ktorom vyberali, ktorý avatar podľa nich dával lepšie rady a mal príjemnejší hlas a hodnotili hlasové charakteristiky avatarov.

Hlasy avatarov sa líšili iba v tempe reči. U prispôsobujúceho sa avatara sa tempo v radách synchronizovalo s tempom hráčových žiadostí o radu. U neprispôsobivého avatara sa buď tempo nemenilo (*Stable*) alebo sa prispôsobovalo asynchronne, teda v opačnom smere (*Antiadapt*). Polovica probantov hrala hru v spôsobe *Adapt-Stable* a druhá polovica v móde *Adapt-Antiadapt*. V obidvoch podskupinách sa polovicí probantov prispôsobovala Monika a polovicí Tereza.

3.3 Analýza dát

Primárne závislé premenné, pomocou ktorých sa testovala základná otázka tohto výskumu, boli dve formy pomeru rád, ktoré si hráč vypýtal od prispôsobujúceho a neprispôsobujúceho avatara. Pri pôvodnom pomere (*Skóre*) sa spočítał počet rád vyžiadaných od jednotlivých avatarov. Pri výhovanom pomere (*Váhované Skóre*) sme sa snažili podchýtiť možnosť, že vplyv prispôsobovania na dôveru hráča k avatarovi sa vyvíja počas hrania a očakávali sme, že vzťah medzi prispôsobovaním a dôveryhodnosťou bude silnejší ku koncu hry ako na začiatku hry. Použili sme jednoduché váhovanie, pri ktorom sa porovnáva súčet poradových čísel ľahov, v ktorých si hráč pýtal radu of jedného alebo druhého avatara. Žiadosti o radu ku koncu hry majú väčšiu váhu ako tie pri začiatku hry.

Štatistická analýza využila párovaný t-test v relevantných podskupinách probantov a tiež aj lineárnu regresiu (*Váhovaného Skóre* ako závislej premennej a prispôsobovania (*Adapt-Stable* a *Adapt-Antiadapt*)

spolu s identitou avatara (*Monika-Tereza*) ako nezávislých premenných.

3.4 Výsledky

Výsledky experimentu ukazujú na vyvrátenie našej základnej hypotézy, že probanti budú viac dôverovať hlasu avatara, ktorý sa im prispôsobuje v jedinej akusticko-prozodickej vlastnosti tempa reči. Párové t-testy v skupine, ktorá hrala hru v móde *Adapt-Antiadapt* preukázali súčasne nesignifikantnú, ale konzistentnú, tendenciu, že hráči si častejšie pýtali radu od toho avatara, ktorý sa im neprispôsoboval (*Antiadapt*), pri neváhovanom skóre to bolo $t[9] = -1,75$, $p = 0,11$ a pri výhovanom $t[9] = -1,68$, $p = 0,13$. Lineárna regresia v tejto skupine ukázala signifikantne negatívny vzťah medzi prispôsobovaním sa a obidvomi verziami *Skóre* v tejto podskupine: $t(17) = -2,41$, $p = 0,028$ pre neváhované a $t(17) = -2,3$, $p = 0,034$ pre výhované. Žiadny signifikantný efekt sme nepozorovali v druhej skupine, ktorá hrala v móde *Adapt-Stable*. Identita avatara (Monika/Tereza) ani pohlavie probanta nemali vplyv na výsledky uvedené vyššie.

3.5 Diskusia

Výsledok, že ľudia viac dôverujú syntetickému hlasu, ktorý sa im neprispôsobuje (t.j. prispôsobuje asymetricky) je v rozpore s predchádzajúcimi výsledkami Levitan (2014), ktoré sú najviac porovnatelné s našou procedúrou, ako aj s očakávaniami, ktoré boli podporené kladným vzťahom medzi rečovým prispôsobovaním a pozitívnymi aspektami ako úspešnosť komunikácie alebo sympathetic spolurečníka, ktorý sa prispôsobuje.

Veľmi podobné výsledky z neváhovaného a výhovaného skóre napovedajú, že v tomto type scenára komunikatívnej hry nedochádza k postupnému upevňovaniu vzťahu medzi rečovým prispôsobovaním a dôveryhodnosťou.

Je možné, že prispôsobovanie iba na základe tempa reči je vnímané inak, ako prispôsobovanie iných prozodických charakteristik ako napríklad výšky hlasu alebo intenzity. V nasledujúcom experimente sme chceli overiť, či súbor charakteristik, podľa ktorých sa syntetický hlas prispôsobuje užívateľovi má vplyv na rozvoj vzťahu medzi rečovým prispôsobovaním a pridali sme k tempu reči aj prispôsobovanie priemernej výšky a intenzity hlasu.

4 Experiment 2

4.1 Probanti

V experimente 2 sa zúčastnilo 24 probantov, z toho 17 žien a 7 mužov opäť väčšinou študentov nitrianskych univerzít.

4.2 Procedúra

Procedúra bola identická ako pri experimente 1 (Podkapitola 3.2), iba spôsob prispôsobovania teraz neobsahoval len tempo reči, ale aj priemernú výšku hlasu (f_0) a intenzitu.

4.3 Analýza dát

Analýza dát bola identická ako pri experimente 1 (Podkapitola 3.3).

4.4 Výsledky

V porovnaní s experimentom 1 sa výsledky Experimentu 2 uberajú opačne: probanti dôverovali viac hlasu avatara, ktorý sa im prispôsoboval ako tomu, ktorý sa im neprispôsoboval. Párové t-testy v skupine, ktorá hrala hru v móde *Adapt-Antiadapt* preukázali signifikantný efekt, že hráči si častejšie pýtali radu od toho avatara, ktorý sa im prispôsoboval (*Adapt*), pri neváhovanom skóre to bolo $t[11] = 1,87$, $p = 0,08$ a pri váhovanom $t[11] = 2,49$, $p = 0,03$. Lineárna regresia v tejto skupine tiež ukázala signifikantne pozitívny vzťah medzi prispôsobovaním sa a obidvomi verziami Skóre v tejto podskupine: $t(21) = 2,58$, $p = 0,017$ pre neváhované a $t(21) = 3,43$, $p = 0,025$ pre váhované. Žiadny signifikantný efekt sme nepozorovali v druhej skupine, ktorá hrala v móde *Adapt-Stable*. Identita avatara (Monika/Tereza) ani pohlavie probanta opäť nemali vplyv na výsledky uvedené vyššie.

4.5 Diskusia

Výsledok tohto experimentu je opačný, ako v Experimente 1. Probanti teraz viac dôverovali tomu avatarovi, ktorého syntetický hlas sa im symetricky prispôsoboval.

Podobne ako v Experimente 1 sa výsledky v neváhovanom a váhovanom skóre líšia len minimálne a ukazujú na robustnosť pozorovania, že vzťah medzi prispôsobovaním sa a dôveryhodnosťou je relatívne stabilný.

	Prispôsobovanie syntetického hlasu v:	Dôveryhodnejší hlas
Exp1	Tempe reči	neprispôsobujúci
Exp2	Tempe, intenzite a výške hlasu	prispôsobujúci

Tab. 1: Sumár výsledkov oboch experimentov

5 Všobecná diskusia

Dlhodobým cieľom tohto výskumu je zistiť, či prispôsobovanie akusticko-prozodických vlastností reči avatarov alebo robotických systémov svojmu používateľovi môže pozitívne ovplyvniť dôveru tohto

používateľa voči samotnému automatickému systému. Úvodné výsledky sumarizované v Tabuľke 1 poskytujú niekoľko zaujímavých zistení.

Za najdôležitejší výsledok považujeme to, že výsledky oboch experimentov ukazujú, že prispôsobovanie v akusticko-prozodických charakteristikách syntetickej reči avatara, alebo iného automatického dialógového systému, má potenciál ovplyvňovať dôveryhodnosť tohto systému v očiach jeho užívateľa.

Zaujímavým a neočakávaným výsledkom je rozdiel v spôsobe, akým je vzťah medzi prispôsobovaním a dôveryhodnosťou ovplyvnený výberom akusticko-prozodických vlastností, ktoré sa na rečovom prispôsobovaní zúčastňujú. Pri prispôsobovaní na v tempe reči sme pozorovali negatívny vzťah - teda že probanti dôverovali prispôsobujúcemu sa avatarovi **menej** ako neprispôsobujúcemu sa - a pri kombinácii tempa reči, výšky a intenzity hlasu sme pozorovali pozitívny vzťah - probanti **viac** dôverovali prispôsobujúcemu sa avatari. Tento výsledok poskytuje niekoľko zaujímavých hypotéz na ďalšie testovanie.

Je možné, že tempo reči je v princípe odlišná charakteristika ako kombinácia výšky tónu a intenzity v tom, ako ovplyvňujú vnímanie aj realizáciu vzťahu užívateľa k automatickému systému. Táto hypotéza je podporená v rámci embodiment kognitívnej vedy v tom zmysle, že výška a intenzita hlasu sú fyziologicky silne prepojené, kde napríklad väčší sub-glótálny tlak má za následok vyššiu intenzitu a výšku základného tónu, a naproti tomu tempo reči je vo veľkej mieri nezávislé od ostatných dvoch charakteristík.

S tým súvisí možnosť, že pozitívne vnemy, ktoré sa môžu spájať so zvýšenou mierou dôveryhodnosti v dialógoch, sa viažu u ľudí na asymetrické charakteristiky tempa reči ale na symetrické charakteristiky výšky a intenzity hlasu. V experimente 1, kde sa prispôsobovalo len tempo reči, tak prevláadol negatívny vzťah a v experimente 2 hlasové charakteristiky prevážili nad tempom reči, čo vyústilo do pozitívneho vzťahu.

Kedže sme v oboch experimentoch pozorovali len minimálne rozdiely medzi váhovaným a neváhovaným skóre, domnievame sa, že to naznačuje stabilitu vzťahu medzi prispôsobovaním a dôveryhodnosťou v tomto scenári komunikatívnej hry.

Je tiež dôležité spomenúť, že preferencia voči prispôsobujúcemu (alebo neprispôsobujúcemu) avatari nemusí automaticky znamenať väčšiu dôveryhodnosť a môže byť spojená s (iba) príjemnejším hlasom toho ktorého avatara. V hre je však hráč motivovaný dosiahnuť čo najlepší výsledok a musí sa spoliehať na radu jedného z avatarov a tak výber avatara (s príjemnejším hlasom) je úzko prepojený na dôveryhodnosť, a my teda skúmame či prispôsobujúci hlas je viac dôveryhodnejší.

V budúcnosti chceme skúmať aj to, či vzťahu medzi rečovým prispôsobovaním a dôveryhodnosťou je

ovplyvnený osobnostnými charakteristikami probanta. Ako bolo popísané v Podkapitole 3.2, máme pre každého probanta dvoch experimentov informácie zo zjednodušeného osobnostného testu a plánujeme preskúmať, či pozitívny a negatívny vzťah medzi prispôsobovaním a dôveryhodnosťou je posilnený alebo oslabený u ľudí s rôznymi osobnostnými charakteristikami.

Podčakovanie

Tento príspevok vznikol s podporou grantovej agentúry Air Force Office for Scientific Research (Air Force Material Command, USAF) v rámci grantu číslo FA9550-15-1-0055 a sčasti aj VEGA v rámci grantu 2/0197/15. Moje podčakovanie patrí aj spolupracovníkom Ramirovi Galvezovi a Agustinovi Gravanovi z University of Buenos Aires, Rivke Levitan z CUNY Brooklyn College v New Yorku, a Marianovi Trnkovi a Milanovi Ruskovi z UI SAV Bratislava.

Literatúra

- Bell, L., Gustafson, J., a Heldner, M. (2003). Prosodic adaptation in human–computer interaction. V zborníku ICPHS, str. 2453–2456.
- Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation* 6(4): 802–813.
- Beňuš, Š., Gravano, A. a Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43(12): 3001–3027.
- Boersma, P. a Weenink, D. (2017). Praat: doing phonetics by computer [computer program]. <http://www.praat.org>.
- Branigan, H., Pickering, M., a Cleland, A. (2000). Syntactic co-ordination in dialogue. *Cognition* 75: B13–B25.
- Branigan, H.P., Pickering, M.J., Pearson, J. a McLean, J.F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics*, 42: 2355–2368.
- Brennan, S. (1996). Lexical entrainment in spontaneous dialog. V zborníku *International Symposium on Spoken Dialog (ISSD)*.
- Giles, H., Coupland, N. a Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. V H. Giles, N. Coupland a J. Coupland *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge: Cambridge University Press, 1–68.
- de Jong, N.H. a Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41: 385.
- Edlund, J., Heldner, M., a Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. V zborníku *Interspeech*.
- Gosling, S., Rentfrow, P. a Swann, W. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality* 37(6): 504–528.
- Levitán, R. (2014). Acoustic-prosodic entrainment in human-human and human-computer dialogue. Ph.D. dissertation, Columbia University.
- Levitán, R., Beňuš, Š., Gravano, A. a Hirschberg, J. (2015). Entrainment in Slovak, Spanish, English, and Chinese: A cross-linguistic comparison. V zborníku *SIGdial*, str. 325–334.
- Levitán, R. a Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. V zborníku *Interspeech*, str. 3081–3084.
- Lubold, N., Walker, E. a H. Pon-Barry. (2016). Effects of Voice-Adaptation and Social Dialogue on Perceptions of a Robotic Learning Companion. V zborníku *ACM/IEEE International Conference on Human Robot Interaction*, str. 255–262.
- Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. V zborníku *Speech Prosody*.
- Nagy, M. (2006). Penguin Quart-Slovak digit speech recognition game based on HMM Marek Nagy. V zborníku *Artificial Intelligence Applications and Innovations*, str. 179–186.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119 (4): 2382–2393.
- Pickering, M. a Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36 (4): 329–347.
- Reitter, D., Keller, F. a Moore, J. 2006. Computational modelling of structural priming in dialogue. V zborníku *HLT/NAACL*.
- Strupka, E., Nieburgh, O. a Fischer, K. (2016). Influence of Robot Gender and Speaker Gender on Prosodic Entrainment in HRI. V zborníku *IEEE International Symposium on Robot and Human Interactive Communication*.
- ten Bosch, L., Oostdijk, N. a Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication* 47: 80–86.

Effect of aging on functional brain networks: a graph-theoretical analysis

Lubica Benuskova¹, Paul McCarthy², Elizabeth A. Franz³

¹Department of Applied Informatics, FMFI, Comenius University, Bratislava, Slovakia

²Department of Computer Science, University of Otago, Dunedin, New Zealand

³Department of Psychology, University of Otago, Dunedin, New Zealand

Email: lubica@ii.fmph.uniba.sk

Abstract

Graph theory is a novel approach to the analysis of functional connectivity in the brain. We have performed graph theoretical analysis of an fMRI data set from a study conducted in 2000 by Buckner et. al. Buckner et al. (2000), in an attempt to find differences between healthy participants, both young and aged, and aged participants with mild Alzheimer's Disease. We apply our techniques to the data, and also give a preliminary statistical comparison across these groups. Our results show that there are statistically significant differences between the groups in several global network characteristics, including density, number of nodes and maximum degree. We hypothesise that these differences reflect a general change in how the aging brain copes with everyday tasks, meaning an increased effort required by aged brains to perform, which is further accentuated by the onset of Alzheimer's disease.

1 Introduction

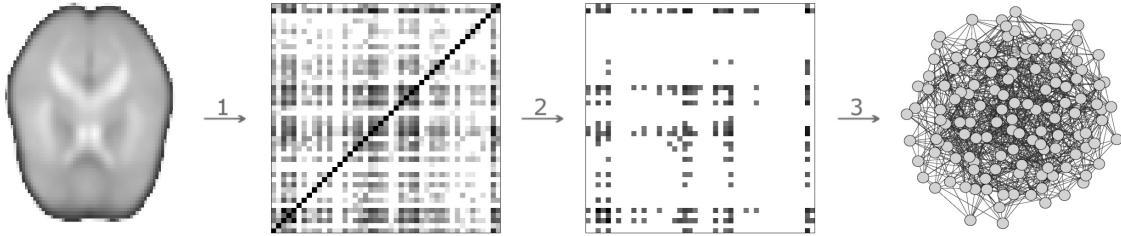
The human brain is an incredibly complex system, consisting of billions of neurons, interacting over trillions of synaptic connections. These interactions give us consciousness and intelligence, control our core functions, and allow us to manipulate our environment. Functional magnetic resonance imaging, or fMRI, allows us to measure neural activity throughout the brain over time. From these measurements, we may gain insights into the functional connectivity of the brain, i.e. how different parts interact with each other when the brain is solving a particular task or just when it is idle. How can we analyse these interactions, and how do they change with age and onset of Alzheimer's disease? This paper attempts to address these questions by using graph theory techniques to analyse functional connectivity derived from fMRI data.

Alzheimer's disease (AD) is the most common form of dementia, causing progressive cognitive decline, with age the primary risk factor (Gauthier (2007)). The pathology of AD is characterised by extracellular deposits of senile, or amyloid plaques, mainly in the cerebral, cortical and hippocampal regions, and neurofibrillary tangles that occupy much of the cyto-

plasm of neurons. In addition, brains from AD patients demonstrate degenerative changes, inflammation as well as neuronal cell loss and synaptic dysfunction, which have been assumed to be consequences of the accumulation of pathological protein components (Gauthier (2007)). Aging by itself causes changes in the brains of healthy people. Global volume of grey matter, i.e. number of neurons, decreases with age, although the effect is not uniformly distributed throughout the brain. It is interesting that global white matter, i.e. number of myelinated axons, does not decline with age, but local areas of relative loss and relative preservation exist.

Functional Magnetic Resonance Imaging, or fMRI, is a technique for capturing high resolution 3D images of oxygenated blood levels in the brain. fMRI measures the ratio of oxygenated to de-oxygenated blood throughout the brain; this ratio is referred to as the Blood Oxygen Level Dependent (BOLD) contrast Ogawa et al. (1992). Evidence suggests that, as neural activity in a brain region increases, a corresponding increase in blood flow to that region occurs, bringing in oxygen to fuel the increased rate of metabolism. Thus, neural activity throughout the brain may be inferred from the subsequent change in the BOLD contrast Ulmer & Jansen (2010); Buxton (2009). A typical fMRI image consists of 3D pixels, known as voxels, each representing an area as small as $\sim 1\text{mm}^3$.

Graph theory has become a popular approach for the analysis of fMRI data (see e.g. Sporns et al. (2004); Bullmore & Sporns (2009); Rubinov & Sporns (2010)). Graph theoretical analysis can provide insights into neural behaviour not readily achievable using more traditional approaches, such as principal component analysis (PCA) Friston et al. (1993), and statistical parametric mapping (SPM) Friston et al. (1990). Friston et. al. Friston et al. (1993) formalised definitions for the three forms of connectivity most commonly of interest: *functional connectivity*, temporal correlation between neural events occurring in spatially distinct brain regions; *effective connectivity*, the influence that one brain region exerts upon another; and *anatomical connectivity*, the physical connections which exist between brain regions. fMRI is an ideal technique for deriving functional and effective connectivity; while anatomical connectivity may be measured using more advanced tech-



Obr. 1: The process of creating a graph from an fMRI data set proceeds as follows: **1.** Temporal correlation values are calculated between the time series data for every pair of voxels, creating a square symmetric correlation matrix. **2.** This matrix is thresholded to remove low correlation values. **3.** The matrix is then converted into a simple undirected, unweighted graph, where nodes represent voxels, and edges represent strong temporal correlation between pairs of voxels.

nologies such as diffusion spectrum or tensor imaging (e.g. Hagmann et al. (2007)), or using statistical methods on MRI data (e.g. He et al. (2007)).

Graphs are an intuitively appealing structure for use in the analysis of functional connectivity: regions of the brain are represented in a graph as nodes, and interaction between regions as edges. A directed graph may be used to model effective connectivity (i.e. causal, or influential relationships between brain regions), and edge weights may be used to represent the strength of inter-regional relationships. In this study however, we have worked only with undirected, unweighted graphs, and are thus assuming that relationships between nodes are symmetrical. The procedure followed in this study is described in Figure 1.

Interpreting the results provided by graph theoretical analysis is a non-trivial task. Rubinov and Sporns Rubinov & Sporns (2010) give an excellent overview of current trends in the interpretation of the various measures that may be calculated upon graphs generated from brain imaging data. These range from simple measures, such as node degree, to more complex high level measures, such as modularity Newman & Girvan (2004) and betweenness centrality Freeman (1979). In this short paper, we will limit our discussion to those measures that are easily interpretable and useful for making inferences about aged related changes in the functional connectivity of the brain; we focus on global differences between the graphs under analysis.

2 Data and Methods

2.1 The data

Structural and functional MRI data, from a study conducted by Buckner et. al. in 2000, were used as a basis for our analysis (Buckner et al. (2000); <http://www.fmridc.org>). Data were acquired from 41 participants: fourteen participants were categorised as ‘young’ (mean age 21.1, S.D. 2.0), fifteen as ‘aged’ (mean age 75.1, S.D. 6.9), and the remaining twelve

(mean age 77.1, S.D. 5.3) as ‘aged with Alzheimer’s Disease (AD)’, having very mild to mild Dementia of the Alzheimer Type according to the Clinical Dementia Rating (CDR) Morris (1997). Seven participants in the aged with AD group had CDR scores of 0.5 (very mild), and the remaining five had CDR scores of 1.0 (mild). There was no statistically significant difference in age between the means of the aged and aged with AD group, as assessed by a two-tailed independent sample *t*-test ($p = 0.58$).

The study involved participants completing a simple visual-motor task, in an event based experimental paradigm Friston et al. (1998). Each participant underwent four fMRI recording sessions, referred to as *runs*. Each run consisted of 128 fMRI images, with a TR (image acquisition) time of 2.68 seconds. During a single run, 15 *trials* were executed, each having a duration of 8 images (21.44 seconds), making a total recording time of approximately 5.5 minutes per run (the first trial in each run began at image #5, and the last trial ended at image #125). A trial consisted of either one or two visual stimuli, presented as a flickering checkerboard pattern; the participants were instructed to push a button with their right index finger upon onset of each stimulus. During a ‘one-stimulus’ trial, the stimulus was triggered at the start of the trial. During a ‘two-stimulus’ trial, the first stimulus was triggered at the start of the trial, and the second stimulus was triggered 5.36 seconds after the first. One- and two-stimulus trials were pseudorandomly inter-mixed.

2.2 Data preprocessing

Data preprocessing consisted of the steps described below. Before any preprocessing, the first four and last four images from every run were discarded, as they were not part of any trial.

1. *Visual inspection:* Every run in the raw data was visually inspected to check for obvious anomalies. This step uncovered three suspect data sets; the data for participant #3 were discarded, due to the

- presence of significant noise throughout every run. The data for participants #15 and #19 contained aliasing effects, which were manually corrected. These three data sets were from the aged group; a two-tailed independent sample *t*-test revealed no age difference between means of the aged and aged with AD group after the removal of participant #3.
2. *Slice-timing correction*: The fMRI slices were captured in an interleaved manner, thus every run was corrected for slice timing differences using Fourier interpolation; this was accomplished with the 3dTshift tool, provided with AFNI¹ (Cox (2011)).
 3. *High-pass temporal filtering*: To correct for scanner drift, and to remove long term trends, a high-pass temporal filter was applied to every run, with a pass frequency of $1/42.88 \sim 0.02\text{Hz}$ (the duration of two trials). This was achieved using the fslmaths utility, provided with FSL² (Smith et al. (2004)).
 4. *Motion correction*: Motion correction was performed using a 6-parameter rigid body transformation. For each run, a mean image was created; every image within the run was then aligned to this mean image. AIR³ was used for motion correction, and for all subsequent image alignment steps Woods et al. (1992).
 5. *Brain segmentation*: Non-brain matter was removed from each participant's structural MRI image, using the bse utility provided with BrainSuite⁴ Shattuck & Leahy (2002); Shattuck et al. (2001).
 6. *Intra-participant registration*: For every run, a mean fMRI image was created, and aligned to the corresponding structural MRI image using a 12-parameter affine transformation. The alignment parameters were then applied to each image in the run, to bring them into alignment.
 7. *Spatial normalisation*: Every image was then aligned to the ICBM452 atlas (Mazziotta et al. (2001)). This was accomplished in two steps: first, the participant's structural MRI image was aligned to the atlas using a non-linear 60 parameter transformation; then, the same alignment parameters were applied to the participant's fMRI images. This two step method is considered to achieve better results than aligning fMRI images directly to the atlas (Filippi (2009); Strother (2006)). Finally,

¹<http://afni.nimh.nih.gov/afni>

²<http://www.fmrib.ox.ac.uk/fsl/>

³<http://bishopw.loni.ucla.edu/air5/>

⁴<http://users.loni.ucla.edu/~shattuck/brainsuite/>

the fMRI images were resampled back to the original resolution of $64 \times 64 \times 16$ voxels.

For each participant, all 60 trial periods were averaged to create one fMRI volume, 21.44 seconds (8 images) in duration; all trials were included in these averages, as it was considered unnecessary to distinguish between 'one-stimulus' and 'two-stimulus' trials Dale & Buckner (1997). Graph theoretical analysis was then applied to these volumes.

2.3 Graph creation

Undirected and unweighted graphs were created from the fMRI data for every participant (see Figure 1). All voxels that exceeded a BOLD level of 200 at any point in time were included as nodes in each graph. The threshold value of 200 was selected after an analysis of the BOLD intensity distribution for each group, as shown in Figure 2. The values below 200 represent non-neural artifacts. Pearson's Correlation Coefficient r (Rodgers & Nicewander (1988)) was calculated between the time series data for all pairs of nodes x and y (Eguiluz et al. (2005)):

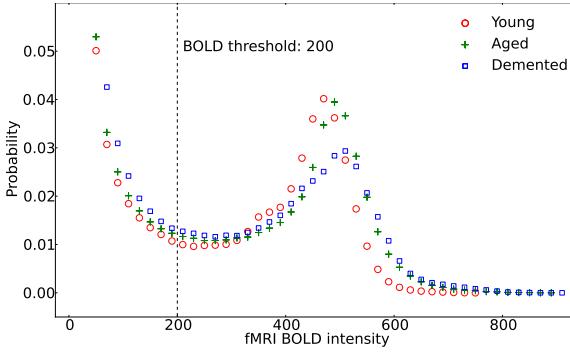
$$r(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \quad (1)$$

where x_i and y_i is the activity in voxels x and y at time i , respectively, and \bar{x} and \bar{y} are the means of the time series data for voxels x and y respectively.

Three graphs for each participant were created from the correlation matrices, using correlation thresholds r_c of 0.8, 0.9 and 0.95. Edges were added between nodes with an absolute correlation value greater than or equal to the threshold; both positive and negative correlations were included to account for excitatory and inhibitory relationships between nodes. Disconnected nodes (nodes which exceeded the BOLD threshold, but were not strongly correlated with any other nodes in the graph), and small components (components less than 20 nodes in size) were removed before further analysis, to prevent skewing of global graph measures. For every graph generated, the remaining nodes formed a single connected component.

2.4 Graph analysis

Standard measures of these undirected, unweighted graphs were calculated, including number of nodes, degree k , density d , degree distribution $P(k)$, maximum degree, characteristic path length L , clustering coefficient C , small-world index, assortativity and global efficiency. The *degree* k of a node is simply the number of neighbours of that node (Diestel (2005)). The *clustering coefficient* C is the ratio of the number of edges which are present between a node's neighbours to the number of possible edges (Watts & Strogatz (1998)). In other



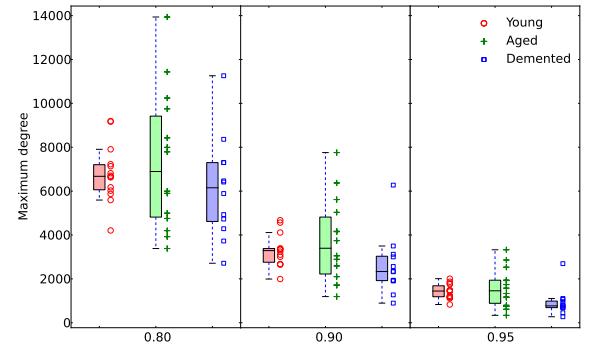
Obr. 2: fMRI BOLD intensity distribution for the young, aged, and aged with AD data sets. A BOLD value of 200 was chosen as the threshold for inclusion of voxels as nodes in the graphs.

words, the clustering coefficient of a node is the density of the subgraph formed by the node's immediate neighbours, and the edges which exist between them. The *density* is the average ratio of true edges to the maximum possible number of edges from each node in the graph. The *characteristic path length L* is the average shortest path length between all pairs of nodes in a graph (Watts & Strogatz (1998)).

The degree distribution $P(k)$ of a graph is the probability distribution of a node in the graph having degree k . A random graph⁵ (P. Erdős (1960)) has a Gaussian degree distribution, with the peak equal to the average node degree. A *scale-free* network, on the other hand, has a power law degree distribution, i.e. $P(k) \sim k^{-\gamma}$, where γ is the power law scaling exponent (Barabasi & Albert (1999)). Scale-free simply means that degrees are not grouped around a single value (peak), or scale, but instead are spread over a wide range of values that can span several orders of magnitude. This implies a potentially rich internal structure, and suggests the presence of highly connected *hubs* in the graph. The maximum degree of a graph therefore gives us the size of the largest hub.

Small-world networks are characterised by a high level of clustering, i.e. $C \gg C_{random}$, combined with a low characteristic path length, i.e. $L \approx L_{random}$ (Watts & Strogatz (1998)). A small-world network is typically portrayed as consisting of densely connected clusters of nodes, with long range connections between clusters. The small-world index (Humphries et al. (2006)) is therefore a combination of two ratios, $\frac{C}{\lambda}$, where $\gamma = \frac{C}{C_{random}}$, and $\lambda = \frac{L}{L_{random}}$. The clustering coefficient of a random graph with number of nodes n and density d is equal to d , and the characteristic path length may be approximated by $\frac{\ln n - 0.5772}{\ln d(n-1)} + \frac{1}{2}$ (Agata Fronczak (2004)). This ratio is used as a measure of 'small-worldness': a graph with a small-world index greater than 1 is considered to have small-world characteristics.

⁵More specifically, an Erdős-Rényi random graph.



Obr. 3: Maximum node degree for the three groups and three correlation thresholds. The mean of aged with AD graphs is significantly lower than young and aged graphs at $r_c = 0.95$ ($p < 0.05$).

Assortativity (Newman (2002)) is a measure of the probability that nodes in a graph are connected to other nodes of similar degree. A graph with a high assortativity indicates that nodes will tend to connect to other nodes which have a similar degree, whereas nodes in a graph with a low (negative) assortativity tend to be connected to nodes with a different degree. *Efficiency* (Vito Latora (2001)) is a measure of information flow through a graph; the global efficiency of a graph is calculated as the average inverse of shortest path lengths between all pairs of nodes. Latora and Marchiori Vito Latora (2001) suggest the use of efficiency as an alternative to the small-world index, to measure the small-world nature of a graph.

3 Results

Group averaged graph measures are provided in Table 1. A two-sample statistical analysis was performed between each pair of groups; Levene's test was used to test for equality of variance, and two-tailed independent sample t -tests (for equal or unequal variances, depending upon the outcome of the Levene's test) were used to test for differences in means, between the three groups. Some interesting results emerge from the figures in Table 1. The main differences which are statistically significant:

- Young graphs have less nodes than both aged and aged with AD graphs (less voxels with a BOLD intensity ≥ 200), significant at $p < 0.001$ for all correlation thresholds. The BOLD intensity distributions in Figure 2 support this: aged and aged with AD participants seem to have more voxels at high intensities than young participants.
- Young graphs have less disconnected nodes (nodes which have no correlation with other nodes) than both aged and aged with AD graphs at $r_c = 0.9$. However, for $r_c = 0.95$ the relationship is reversed:

Tab. 1: Mean (standard deviation in brackets) graph measures for the three groups, at each correlation threshold.

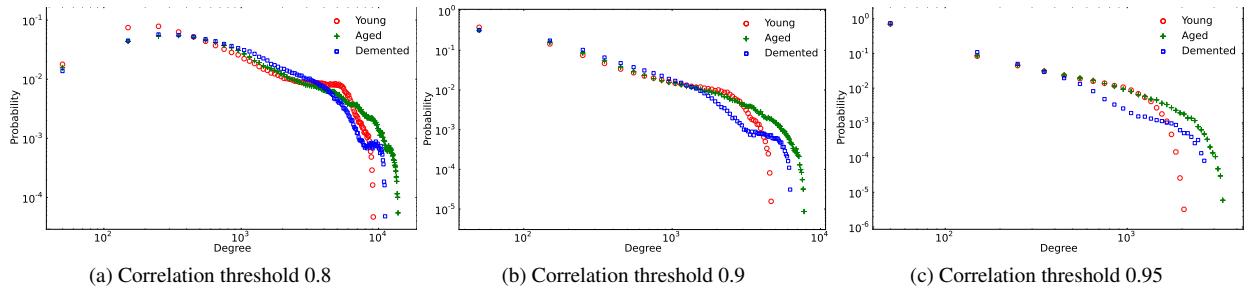
	Young		Aged		Aged with AD	
Correlation threshold 0.8						
<i>Disconnected nodes</i> ^a	0.0	(0.0)	0.0	(0.0)	0.0	(0.0)
Nodes	22888.6	***§§§(996.1)	24750.1	*** (1287.9)	24298.8	§§§ (864.0)
Density	0.0937	(0.0302) ^{**}	0.1064	(0.0604) ^{**}	0.0813	(0.0406)
Degree	2132.65	(662.65) ^{**}	2618.53	(1502.60) ^{**}	1972.83	(987.07)
Maximum degree	6806.43	(1328.60) ^{**}	7337.86	(3157.08) ^{**}	6114.42	(2306.98)
Path length	2.3620	(0.0853) ^{**}	2.3336	(0.1606) ^{**}	2.3824	(0.1108)
Clustering coefficient	0.5012	(0.0435) ^{**}	0.5135	(0.0826) ^{**}	0.4823	(0.0608)
Small-world index	4.2068	(0.9202) ^{**}	4.5307	(1.8788) ^{**}	5.0475	(1.5562)
Assortativity	0.5371	(0.0749)	0.5273	(0.1215)	0.5440	(0.0690)
Global efficiency	0.4709	(0.0242) ^{**}	0.4799	(0.0469) ^{**}	0.4634	(0.0318)
Correlation threshold 0.9						
<i>Disconnected nodes</i> ^a	1.50 ^{**§}	(3.37) ^{**§}	5.57 ^{**}	(4.65) ^{**}	5.67 [§]	(6.02) [§]
Nodes	22887.1	***§§§(995.9)	24744.6	*** (1289.4)	24293.1	§§§ (867.4)
Density	0.0263	(0.0109) [*]	0.0316	(0.0225) [*]	0.0206	(0.0157)
Degree	597.45	(241.11) ^{**}	775.40	(561.58) ^{**}	499.64	(383.06)
Maximum degree	3294.36	(745.07) ^{**§}	3720.36	(1931.91) ^{**}	2588.33	(1379.51) [§]
Path length	3.4689	(0.1582) ^{**§}	3.3869	(0.2913) ^{**}	3.4713	(0.2090) [§]
Clustering coefficient	0.4253	(0.0332) ^{**}	0.4355	(0.0629) ^{**}	0.4119	(0.0448)
Small-world index	10.50 [§]	(3.47) ^{***}	12.99	(8.45) ^{***}	16.08 [§]	(7.67)
Assortativity	0.5969	(0.0657) [*]	0.5901	(0.1120) ^{*†}	0.6275	(0.0692) [†]
Global efficiency	0.3279	(0.0207) ^{**}	0.3404	(0.0414) ^{**}	0.3252	(0.0292)
Correlation threshold 0.95						
<i>Disconnected nodes</i> ^a	764.79 ^{*§§}	(126.62)	647.86 [*]	(118.03)	629.83 ^{§§}	(133.20)
Nodes	22123.8	***§§§(1084.8)	24102.3	*** (1333.3)	23668.9	§§§ (836.6)
Density	0.0069	(0.0032) [*]	0.0081	(0.0064) [*]	0.0047	(0.0046)
Degree	151.20	(67.58) ^{**}	194.34	(155.53) ^{**}	110.15	(109.57)
Maximum degree	1425.79 ^{§§}	(336.23) ^{**}	1575.58 [†]	(879.49) ^{**}	916.75 ^{§§†}	(606.22)
Path length	5.7920	(0.4651) ^{**}	5.5162	(0.7766) ^{**}	5.6872	(0.5816)
Clustering coefficient	0.3933	(0.0220) ^{**}	0.3917	(0.0425) ^{**}	0.3787	(0.0327)
Small-world index	28.12 [§]	(11.87) ^{***§}	43.27	(37.10) ^{***}	60.18 [§]	(38.09) [§]
Assortativity	0.6555 [§]	(0.0495) [*]	0.6504	(0.0849) [*]	0.7035 [§]	(0.0654)
Global efficiency	0.2100	(0.0221) ^{**}	0.2196	(0.0399) ^{**}	0.2064	(0.0284)

^a These values include both disconnected nodes and components less than 20 nodes in size. These nodes were removed before any further processing.

* Significantly different between young and aged (* $p < 0.05$, ** $p < 0.015$, *** $p < 0.001$).

§ Significantly different between young and aged with AD (§ $p < 0.05$, §§ $p < 0.015$, §§§ $p < 0.001$).

† Significantly different between aged and aged with AD († $p < 0.05$, †† $p < 0.015$, ††† $p < 0.001$).



Obr. 4: Accumulated degree distribution for the three groups at each correlation threshold.

young graphs have more disconnected nodes than both aged and aged with AD graphs at the higher threshold.

- Aged with AD graphs have a greater small-world index than young graphs, significant at $p < 0.05$ for $r_c \geq 0.9$.
- The only difference between the aged and aged with AD groups is the maximum degree at correlation threshold $r_c = 0.95$: aged with AD graphs have a lower maximum degree than both young and aged graphs. By looking at the boxplot (Figure 3), we can see that this trend exists at the other thresholds, but is only statistically significant at $r_c = 0.95$.
- Assortativity of the aged with AD graphs is higher than in the young graphs ($p < 0.05$), at $r_c = 0.95$.
- Variances of almost all topological characteristics are significantly higher for aged graphs, when compared to young graphs. There are less differences in variance between the young and aged with AD graphs.

4 Discussion

4.1 AD, but not age, leads to reduced functional connectivity

Despite no significant difference in average degree or density between any of the three groups, aged with AD graphs have a lower maximum degree than both young and aged graphs at correlation threshold $r_c = 0.95$. This seems to contradict the trend portrayed by the accumulated degree distributions, shown in Figure 4. However, further exploration reveals that the aged with AD data are heavily skewed by a single graph with a much greater density - denser than any other graph in the group by a factor of at least 1.6 at $r_c = 0.8$, and at least 2.2 at $r_c \geq 0.9$. This same outlier is evident in Figure 3, for $r_c \geq 0.9$. Figure 5 displays the accumulated degree distribution for $r_c = 0.95$, with this outlier graph removed. In fact, by removing this graph from the data set, and repeating the two-tailed independent sample t -tests, we find that the aged with AD group has significantly lower density, degree and maximum degree than both the young and aged groups, for $r_c \geq 0.9$, as shown in Table 2. The removed subject had a CDR score of 0.5, it is plausible that this was a case of misdiagnosis.

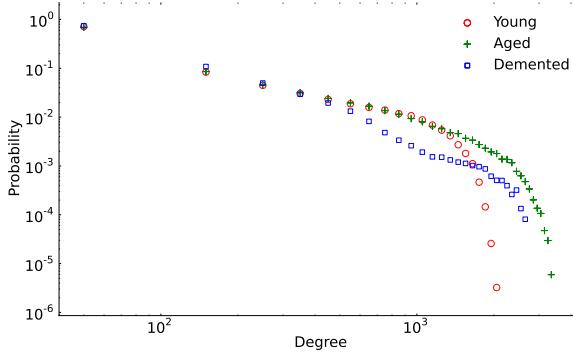
Pathologically, these results make more sense: the neuronal and synaptic atrophy which is known to occur in AD should have a marked impact on the functional connectivity of the brain. These findings are also consistent with other studies on the effects of AD on functional connectivity: Leuchter et. al. Leuchter et al. (1992) found consistently reduced coherence at all

frequency bands, in AD subjects, using EEG; Greicius et. al. Greicius et al. (2004), working with the data set from Buckner et. al.'s study Buckner et al. (2000) (the same data set that was used in this study), found disrupted connectivity around the posterior cingulate and hippocampus in the aged with AD group, using independent component analysis (ICA) (Beckmann & Smith (2004); Sorg et. al. Sorg et al. (2007)) found reduced functional connectivity between the hippocampus and temporal lobe, in subjects at risk of developing AD, using ICA on resting-state fMRI data; Wang et. al. Wang et al. (2007) found reduced inter-lobe connectivity patterns, but increased intra-lobe connectivity, in early AD patients, by calculating and comparing correlation coefficients between all pairs of regions; Supekar et. al. (Supekar et al. (2008)) reported reduced regional connectivity in resting-state fMRI data acquired from AD patients, using wavelet analysis and graph theory.

The increased connectivity present in the aged group relative to the young group, clearly visible in the degree distribution plots (Figures 4 and 5) is a matter of some interest; this difference is still present when the degree values are normalised by graph size, to degree centrality (Freeman (1979)). It is known that the distribution of neural metabolism throughout the brain changes with age, giving credence to the idea that we develop compensatory and adaptive mechanisms to overcome the pathological changes associated with aging. However, studies to date on the functional connectivity of aging (e.g. Greicius et al. (2004); Stam et al. (2007, 2009); de Haan et al. (2009); Buckner et al. (2009)) have found reduced, or non-uniform changes in the functional connectivity of aged subjects, when compared with young subjects. This discrepancy may be caused by the reduced signal-to-noise ratio in fMRI BOLD signals from aged brains (van den Heuvel et al. (2009)); an effect which could also explain the higher variances exhibited in the aged group for many of the network measures. Comparison between these studies is somewhat difficult, due to methodological differences, however, the increased functional connectivity and neural metabolism in the aged group, observed in this study, supports our hypothesis that aged brains have to work harder to perform; this is discussed further in the next section.

4.2 Aged brains work harder

Young graphs have less nodes than both aged and aged with AD graphs (i.e. less voxels with a BOLD intensity ≥ 200), significant at $p < 0.001$. An increased BOLD signal means an increase in oxygenated blood flow, which implies increased neural metabolism. Therefore this finding may mean that the aged and aged with AD participants had to work harder than the young participants to perform the task. This finding is in agreement with findings of increased frontal activation in aged



Obr. 5: Accumulated degree distribution for the three groups, at correlation threshold $r_c = 0.95$. A single outlier graph has been removed from the aged with AD group, dramatically affecting the distribution.

brains (He et al. (2008); de Haan et al. (2009); Buckner et al. (2009)). An alternate or additional explanation is skewing of the BOLD distribution by increased ventricular, and hence cerebrospinal fluid volume, in the aged and aged with AD groups (Buckner et al. (2009)).

We found opposing trends for disconnected nodes at correlation thresholds 0.9 and 0.95. For $r_c = 0.9$, young graphs had less disconnected nodes than both aged and aged with AD graphs. That is, less nodes in the aged and aged with AD graphs than in the young graphs were functionally connected with any other nodes at a correlation of 0.9. The total number of disconnected nodes in each group, however, was less than 0.001% of the total number of nodes. For $r_c = 0.95$, young graphs had more disconnected nodes than both aged and aged with AD graphs. The total number of disconnected nodes in each group at $r_c = 0.95$ was again quite small, at around 5%, so this could be considered to be nothing more than noise.

4.3 Strong small-world characteristics

An interesting finding was that aged with AD graphs have a larger small-world index than young graphs for thresholds $r_c \geq 0.9$ (this difference was still significant at $p < 0.05$ after removal of the outlier from the aged with AD group at both correlation thresholds⁶). There was no significant difference between the aged and aged with AD groups, possibly due to the large variances in these two groups. All three groups exhibit extremely high clustering when compared to random graphs of the same density. In particular, despite having lower densities, the aged with AD graphs maintain their level of clustering at the high correlation threshold of 0.95. The high clustering in each group is enough to overwhelm the fact that, at all correlation thresholds, the characteristic path length is higher than that of an equivalent

⁶This outlier is not the same subject as the dense outlier previously discussed.

random graph.

This finding is in contrast to a number of studies (e.g. Stam et al. (2007, 2009); Supekar et al. (2008); de Haan et al. (2009)), which all found decreased clustering and increased path length in AD subjects, and a corresponding reduction in the small-world index. However, these studies are not directly comparable due to methodological differences. A common approach in graph theoretical analysis is to vary the correlation (or analogous measure) threshold for each graph, ensuring that all graphs to be compared are of similar densities. Our approach is different, in that by using the same threshold for each group, we are comparing graphs of varying densities, but at the *same level of functional connectivity*; we feel this is an important characteristic to retain, when comparing across groups. Furthermore, as no differences were found between the three groups in clustering or path length, we argue that the difference in magnitude of the small-world index (which is wholly derived from these parameters and graph density) primarily reflects the difference in densities between the graphs. This does not discount the fact that all graphs in the study show a strong small-world structure and internal clustering, which implies the presence of a very strong modular structure, and alone warrants further regional analysis of the data.

The accumulated degree distributions shown in Figures 4 and 5 closely resemble that of a decaying, or truncated power-law, as described by Amaral et al. Amaral et al. (2000), and reported in a number of other studies on functional connectivity (e.g. Eguiluz et al. (2005); Bassett & Bullmore (2006); Achard et al. (2006)). This lends weight to the idea that the ‘preferential attachment’ model Barabasi & Albert (1999) is not suitable for modelling real world biological networks; once a node reaches a certain connectivity, there is a *physical cost* associated with acquiring more links Achard et al. (2006); Bullmore & Sporns (2009). Buckner et al. Buckner et al. (2009) found that high-degree hubs in the cerebral cortex are more likely to be affected by amyloid deposition than their low-degree counterparts, giving a plausible explanation for the lower connectivity exhibited by the aged with AD group (after removal of the dense outlier graph).

5 Conclusion

A potentially revealing finding is the fewer nodes found in the young controls compared to the older groups. This points to age as a mediating factor in the work required to perform the task, a possibility that is corroborated by our findings on the increased functional connectivity in aged brains. There was strong evidence of small-worldness in all groups, and particularly high clustering even in the aged with AD group, suggesting the presence of a highly modular structure. Re-

Tab. 2: Mean (standard deviation in brackets) graph measures, after removal of the outlier graph from the aged with AD group.

Aged with AD		
Correlation threshold 0.9		
Density	0.0165 ^{§†}	(0.0072) ^{††}
Degree	398.90 ^{§†}	(165.72) ^{††}
Maximum degree	2253.27 ^{§§†}	(781.96) ^{††}
Correlation threshold 0.95		
Density	0.0034 ^{§§†}	(0.0017) ^{††}
Degree	80.29 ^{§§†}	(37.98) ^{††}
Maximum degree	755.18 ^{§§§††}	(244.29) ^{††}

[§] Significantly different between young and aged with AD
(§ $p < 0.05$, §§ $p < 0.015$, §§§ $p < 0.001$).

[†] Significantly different between aged and aged with AD
(† $p < 0.05$, †† $p < 0.015$, ††† $p < 0.001$).

duced density in the aged with AD group is suggestive of reduced functional connectivity, related specifically to disease and not age alone, and closer analysis of one rather extreme case in the aged with AD group confirms the volatility of the disease on measures of functional connectivity.

References

- Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.*, **26**(1), 63–72.
- Agata Fronczak, Piotr Fronczak, J. A. H. (2004). Average path length in random networks. *Physical Review E*, **70**, 056110.
- Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of Small World Networks. *Proceedings of the National Academy of Sciences*, **97**(21), 11149–11152.
- Barabasi, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, **286**, 509–512.
- Bassett, D. S. & Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, **12**(6), 512–23.
- Beckmann, C. F. & Smith, S. M. (2004). Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, **23**(2), 137–152.
- Buckner, R. L., Snyder, A. Z., Sanders, A. L., Raichle, M. E., & Morris, J. C. (2000). Functional Brain Imaging of Young, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, **12**(Supplement 2), 24–34.
- Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., Andrews-Hanna, J. R., Sperling, R. A., & Johnson, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer’s disease. *J. Neurosci.*, **29**(6), 1860–73.
- Bullmore, E. & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, **10**(3), 186–98.
- Buxton, R. B. (2009). *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, Second edition.
- Cox, R. W. (2011). AFNI: What a long strange trip it’s been. *Neuroimage*, **In Press**.
- Dale, A. M. & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, **5**(5), 329–340.
- de Haan, W., Pijnenburg, Y. A. L., Strijers, R. L. M., van der Made, Y., van der Flier, W. M., Scheltens, P., & Stam, C. J. (2009). Functional neural network analysis in frontotemporal dementia and Alzheimer’s disease using EEG and graph theory. *BMC Neurosci*, **10**, 101.
- Diestel, R. (2005). *Graph Theory*. Springer-Verlag, 3rd edition.
- Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. (2005). Scale-Free Brain Functional Networks. *Physical Review Letters*, **94**(1), 018102–1 – 018102–4e.
- Filippi, M., editor (2009). *fMRI Techniques and Protocols*, volume 41 of *Neuromethods*. Humana Press.
- Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks*, **1**(3), 215–239.
- Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., & Frackowiak, R. S. J. (1990). The Relationship Between Global and Local Changes in PET Scanas. *Journal of Cerebral Blood Flow and Metabolism*, **10**, 458–466.
- Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. J. (1993). Functional Connectivity: The Principal-Component Analysis of Large (PET) Data Sets. *Journal of Cerebral Blood Flow and Metabolism*, **13**(1), 5–14.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-Related fMRI: Characterizing Differential Responses. *NeuroImage*, **7**(1), 30–40.

- Gauthier, S., editor (2007). *Clinical Diagnosis and Management of Alzheimer's Disease*. Informa Healthcare, Third edition.
- Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *PNAS*, **101**(13), 4637–4642.
- Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., & Thiran, J.-P. (2007). Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE*, **2**(7), e597.
- He, Y., Chen, Z. J., & Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex*, **17**(10), 2407–19.
- He, Y., Chen, Z., & Evans, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *J. Neurosci.*, **28**(18), 4756–66.
- Humphries, M. D., Gurney, K., & Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proc. Biol. Sci.*, **273**(1585), 503–11.
- Leuchter, A. F., Newton, T. F., Cook, I. A., Walter, D. O., Rosenberg-Thompson, S., & Lachenbruch, P. A. (1992). Changes in Brain Functional Connectivity in Alzheimer-Type and Multi-Infarct Dementia. *Brain*, **115**(5), 1543–1561.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., & Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **356**(1412), 1293–322.
- Morris, J. C. (1997). Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*, **9**(S1), 173–176.
- Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters*, **89**, 208701.
- Newman, M. E. J. & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, **69**, 026113.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, **89**(13), 5951–5955.
- P. Erdős, A. R. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 17–61.
- Rodgers, J. L. & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, **42**(1), 59–66.
- Rubinov, M. & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, **52**(3), 1059–69.
- Shattuck, D. W. & Leahy, R. M. (2002). BrainSuite: An automated cortical surface identification tool. *Medical Image Analysis*, **6**(2), 129–142.
- Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., & Leahy, R. M. (2001). Magnetic Resonance Image Tissue Classification Using a Partial Volume Model. *NeuroImage*, **13**(5), 856–876.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobniak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, **23 Suppl 1**, S208–19.
- Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., Drzezga, A., Förstl, H., Kurz, A., Zimmer, C., & Wohlschläger, A. M. (2007). Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proceedings of the National Academy of Sciences*, **104**(47), 18760–18765.
- Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci. (Regul. Ed.)*, **8**(9), 418–25.
- Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., & Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cereb. Cortex*, **17**(1), 92–9.
- Stam, C. J., de Haan, W., Daffertshofer, A., Jones, B. F., Manshanden, I., van Cappellen van Walsum, A. M., Montez, T., Verbunt, J. P. A., de Munck, J. C., van Dijk, B. W., Berendse, H. W., & Scheltens, P. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease. *Brain*, **132**(Pt 1), 213–24.

Strother, S. C. (2006). Evaluating fMRI Preprocessing Pipelines. *IEEE Engineering in Medicine and Biology Magazine*, **25**(2), 27–41.

Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput. Biol.*, **4**(6), e1000100.

Ulmer, S. & Jansen, O., editors (2010). *fMRI: Basics and Clinical Applications*. Springer.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Efficiency of functional brain networks and intellectual performance. *J. Neurosci.*, **29**(23), 7619–24.

Vito Latora, M. M. (2001). Efficient Behavior of small-world networks. *Physical Review Letters*, **87**(19), 198701.

Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., & Jiang, T. (2007). Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Hum Brain Mapp*, **28**(10), 967–78.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, **393**, 440–442.

Woods, R., Cherry, S., & Mazziotta, J. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, **16**(4), 620–633.

Etické a legislatívne problémy využívania dronov

Ivana Budinská

Ústav informatiky Slovenskej akadémie vied
Dúbravská cesta 9, 845 07 Bratislava, Slovensko
Email: budinska@savba.sk

Abstrakt

Súčasný rozvoj informačných technológií vyvoláva nové otázky týkajúce sa etiky a morálky vo vzťahu k inteligentným a čoraz viac autonómnym systémom. Robotika patrí k tým oblastiam, ktorých rozvoj je priamo podmienený rozvojom informačných technológií a umelej inteligencie. Dnes hovoríme o štvrtej generácii robotov, ktoré sú vysoko autonómne a pri plnení svojich úloh často prichádzajú do kontaktu s ľuďmi, resp. priamo s nimi kooperujú. Konštruktéri a vývojári takýchto robotických systémov musia predovšetkým riešiť otázky bezpečnosti ľudí. Nemenej dôležité sú však aj legislatívne a etické otázky, na ktoré sa musí myslieť už vo fáze návrhu a vývoja inteligentných a autonómnych robotických systémov. Tento príspevok sa orientuje predovšetkým na etické a legislatívne problémy súvisiace s prevádzkou bezpilotných vzdušných prostriedkov.

1 Úvod

Roboetika je pomerne nová vedná disciplína, ktorej vznik si vynútil prudký rozvoj robotických systémov a ich využívanie v bežnom živote. V literatúre sa často stretávame s definíciou štyroch generácií robotov. V prevej generácii robotov ide o mechanické programovateľné zariadenia, ktoré sa využívajú v priemysle a do kontaktu s nimi prichádzajú zaškolení pracovníci. Roboty druhej generácie sa objavujú v 80-tych rokoch minulého storočia. Sú vybavené senzorickými systémami, výpočtovým výkonom a určitým stupňom strojovej inteligencie. To im umožňuje vykonávať synchronizované úkony aj bez sústavného dohľadu človeka. Ich pôsobnosť je však stále obmedzená na špecifické priestory priemyselnej prevádzky a neprichádzajú do kontaktu s bežnými ľuďmi. Až roboty tretej a štvrtnej generácie sú schopné kooperovať s ľuďmi a pracovať aj v bežných prostrediaciach. Čoraz častejšie sa stretávame s robotmi v oblasti starostlivosti o hendikepovaných a starších ľudí, ale aj v iných službách. Experimentuje sa s robotickými čašníkmi, drony sa využívajú na doručovanie zásielok. Dôležitú úlohu zohrávajú pri záchrannárskych prácach ale aj pri voľnočasových aktitvách. Nárast výpočtovej kapacity, ktorú môžu roboty využívať, ako aj rozvoj metód umelej inteligencie a

kognitívnych schopností robotických systémov, stavajú pred nás nové etické, morálne a legislatívne problémy, s ktorými sa tradičné chápanie etiky a morálky nevie vysporiadať. Musíme sa pripraviť na odpovedanie otázok týkajúcich sa zodpovednosti za konanie autonómnych robotov, otázok autorských práv, sociálneho vplyvu na spoločnosť a mnohé ďalšie. Spisovateľ Isaac Asimov už v roku 1942 v poviedke Run-Around sformuloval tri zákony robotiky, ktoré stále tvoria základ nášho chápania novej etiky vo vzťahu k robotickým systémom:

1. Robot nesmie zraniť človeka, ani svojím nekonaním dopustiť, aby bol človek zranený.
2. Robot musí poslúchať príkazy dané človekom s výnimkou prípadov, keď takýto príkaz je v rozpore so zákonom číslo 1.
3. Robot musí chrániť vlastnú existenciu dovtedy, kým nedôjde ku konfliktu so zákonom číslo 1 alebo 2.

Ide však o veľmi zjednodušený pohľad na problematiku etiky v robotike. Keďže robotika integruje výskum a výstupy ďalších súvisiacich oblastí - napr. informatiky, strojárstva, elektrotechniky - aj etika pre robotiku vychádza z etických princípov platných pre tieto oblasti. Predovšetkým etický kódex pre informatiku známy pod skratkou PAPA (z anglického Privacy, Accuracy, Property, Accessibility) a kódex inžinierskej etiky, ktorý sa zaobera najmä otázkami bezpečnosti a spoľahlivosti systémov, sú pre roboetiku relevantné. Problematika etiky v robotike je veľmi široká. Preto sa v článku sústredíme na otázky a problémy týkajúce sa prevádzky bezpilotných vzdušných prostriedkov - UAV (z angl. Unmanned Airborne Vehicles). V nasledujúcej kapitole sú zadefinované zariadenia UAV a sú vymedzené základné okruhy problémov legislatívnych a etických. V kapitole 3 sa nachádza prehľad aktuálneho stavu legislatívny pre UAV na Slovensku v kontexte práva Európskej únie. Kapitola 4 nastoľuje špecifické etické problémy v súvislosti so širokým využívaním UAV. Kapitola 5 je venovaná otázkam využívania UAV pre vojenské a obranné účely. V záveru článku je načrtnutý ďalší rozvoj roboetiky, ako vednej disciplíny, jej význam a jej aplikácie pre oblasť bezpilotných vzdušných prostriedkov.

2 Bezpilotné vzdušné prostriedky

Pod pojmom bezpilotný vzdušný prostriedok, často nazývaný aj dron, rozumieme diaľkovo riadené alebo autonómne lietadlo alebo model lietadla, resp. vzdušného prostriedku. Telo dronu môže mať rôzne podoby a rôzne formy pohonov. Najčastejšie sa stretávame s dronmi, ktoré sa pohybujú za pomoci vrtuľ. Podľa počtu vrtuľ rozpoznávame helikoptéry, trikoptéry, kvadkoptéry, multikoptéry. Drony však môžu mať aj inú konštrukciu, napríklad vo forme svahového krídla. Súčasťou UAV je aj nezávislý systém riadenia, ktorý je umiestnený priamo na zariadení a umožňuje autonómne vykonávanie niektorých úloh. Drony sú vybavené výkonnými senzormi, najmä kamerovými systémami a rôznymi druhami snímačov na rozpoznávanie prostredia a objektov v prostredí. Okrem toho môže byť dron vybavený aj nosným zariadením na prevádzanie ďalších objektov podľa konkrétneho spôsobu využitia. Samotný prostriedok je súčasťou širšieho systému, ktorý pozostáva z ďalších súčasťí. K najdôležitejším patrí pozemný, resp. vzdialený systém riadenia, komunikácia, systém na prenos, uchovávanie a spracovanie dát a ďalšie podporné systémy. Do kategórie lietadiel schopných lietať bez pilota patrí veľké množstvo zariadení, od lietadiel, ktoré sú veľkosťou a vybavením porovnateľné s pilotovanými lietadlami až po veľmi malé lietajúce zariadenia používané ako hračky alebo vo výskume. Z hľadiska veľkosti delíme drony do štyroch kategórií:

- nano a mikro drony, ktoré sa používajú predovšetkým vo výskume ale predpokladá sa ich využívanie napríklad aj v medicíne,
- malé drony, ktorých veľkosť sa pohybuje v desiatkach centimetrov a v súčasnosti sa využívajú najmä na rekreačné a športové účely,
- stredné drony, ktoré sú určené najmä na komerčné použitie,
- veľké a veľmi veľké dróny, ktoré sa svojou veľkosťou približujú k veľkým pilotovaným objektom a ich používanie na civilné účely sa zatiaľ nepredpokladá.

Najmä malé drony sú v súčasnosti veľmi rozšírené. Je to spôsobené ich cenovou dostupnosťou a jednoduchou obsluhou.

Ďalší spôsob kategorizácie dronov je na základe ich výkonu, t.j. na základe veľkosti priestoru, v ktorom dokážu operovať. Tá je daná jednak senzorickým vybavením a jednak výkonnosťou batérií.

Použitie dronov je veľmi rozsiahle a stále sa ukazujú nové možnosti. Vo všeobecnosti môžeme kategorizovať tri typy použitia dronov: rekreačné, komerčné a vojenské. O etických problémoch pre jednotlivé typy použitia sa píše v článku Wilson (2014)

Rekreačné používanie dronov dnes zahŕňa okrem leteckých modelárov aj nadšencov z radov fotografov a filmárov. Medzi komerčné používanie zaradujeme využívanie UAV na účely monitorovania a stráženia objektov, monitorovania dopravnej situácie, na kontrolu diaľkových infraštruktúr (elektrické vedenia, plynovody, ropovody), na získavanie informácií pre predpovedanie počasia. Veľký potenciál má využitie dronov v poľnohospodárstve a lesníctve, ale aj v službách, napríklad na doručovanie zásielok. Stále väčší význam budú mať drony pri záchranárskych prácach na prehľadávanie postihnutých oblastí a na vyhľadávanie obetí. Robia sa experimenty pri ktorých drony dokážu poskytnúť postihnutým ľuďom nevyhnutnú prvú pomoc a prepraviť ich do bezpečia.

Drony pracujú samostatne alebo v skupine. Aktuálnym trendom vo výskume je koordinácia skupiny robotov a to aj heterogénnej, t.j. takej, kde sa vyskytujú vzdušné a pozemné roboty, prípadne aj roboty pohybujúce sa vo vodnom prostredí.

Aplikačné možnosti bezpilotných vzdušných prostriedkov sú veľmi veľké. Ich široké využívanie kladie nároky na legislatívne a etické otázky, ktoré musíme riešiť. O súčasnom stave legislatívy a niektorých problémoch s jej uplatňovaním sa píše v nasledujúcej kapitole.

3 Súčasný stav legislatívy pre prevádzku UAV

Prevádzkovanie bezpilotných vzdušných prostriedkov zasahuje do viacerých oblastí bezpečnosti a ochrany obyvateľstva. Európska únia zriadila Európsku agentúru pre bezpečnosť letectva - EASA (z angl. European Aviation Safety Agency), ktorej úlohou je koordinovať postup štátov Európskej únie v oblasti civilného letectva.

EASA v dokumente EASA (2015) definuje tri kategórie bezpečnostných požiadaviek na drony:

- otvorená kategória - nízke riziko: nie sú potrebné povolenia na prevádzku mimo zakázaných oblastí. Tie sú definované národnými orgánmi. Bezpečnosť zaručujú bezpečnostné predpisy pre daný produkt a minimálny súbor operačných pravidiel, ako aj obmedzenia na prevádzkovú hmotnosť, rýchlosť a spôsob komunikácie.
- špecifická kategória - stredné riziko: vyžaduje sa povolenie národným orgánom, ktoré sa vydáva na základe posúdenia rizík pre konkrétné zariadenie. EASA poskytuje zoznam niektorých štandardov a parametrov ako pomôcku pre posúdenie miery rizika.
- autorizovaná kategória - vysoké riziko: požiadavky sú porovnateľné s požiadavkami na pilotované vzdušné prostriedky.

Obmedzenia prevádzky UAV sa netýkajú len bezpečnosti. V prípade, že sú UAV vybavené prostriedkami pre záznam (obrazový, zvukový a pod.), je potrebné regulovať aj ich používanie vzhľadom na ochranu súkromia. UAV sa čoraz častejšie požívajú na monitorovanie hromadných akcií, športových a kultúrnych podujatí a na rôzne druhy leteckých predstavení. V niektorých štátach sa na tieto účely vyžaduje poistenie.

3.1 Platná legislatíva pre UAV na Slovensku

Na Slovensku v súčasnosti prevádzku bezpilotných vzdušných prostriedkov upravuje rozhodnutie Dopravného úradu č. 1/2015 z 19. augusta 2015 DÚ/L001-A/v3 (2015), ktorým sa určujú podmienky vykonania letu lietadlom spôsobilým lietať bez pilota vo vzdušnom priestore Slovenskej republiky. Táto vyhláška definuje bezpilotné vzdušné prostriedky, na ktoré sa vzťahujú jej ustanovenia, definuje osoby, ktoré môžu byť držiteľmi povolenia na lietanie s UAV a zodpovednosť za letovú spôsobilosť takého zariadenia. Vyhláška sa nevzťahuje na šarkany a neobsadené balóny. Táto vyhláška upravuje podmienky vykonania letu autonómny a diaľkovo riadeným lietadlom, technické parametre takýchto prostriedkov a požiadavky na vykonávanie letu v riadenom vzdušnom priestore. Vykonanie letu autonómnym lietadlom je vo vzdušnom priestore zakázané. Rozhodujúcim kritériom pre vykonávanie letov je podľa vyhlášky maximálna vzletová hmotnosť. Prostriedky s maximálnou vzletovou hmotnosťou 20 kg a vyššou podliehajú evidencii na Dopravnom úrade. Kategorizácia len na základe hmotnosti však nie je dostačujúca, pretože je zrejmé, že aj kinetický výkon môže byť rozhodujúci pre potenciálne ohrozenie osôb a vecí. Vyhláška zavádzajúca aj ďalšie povinnosti pre osoby, ktoré ovládajú bezpilotné lietadlá. Okrem iného aj povinnosť viesť záznamy o vykonaných letoch v palubnom denníku. Ďalšie obmedzenia sú v súvislosti s leteckým snímkováním. Ak sa bezpilotné prostriedky používajú na snímkovanie, podliehajú osobitnému predpisu Leteckého zákona a nesmú sa prevádzkovať bez povolenia. Samotná existencia záznamového zariadenia na UAV však nezakladá povinnosť žiadať o povolenie. Situácia sa zmení, ak sa toto zariadenie začne používať. Vykonávanie leteckého snímania bez povolenia je sankcionované. Tu je potrebné pripomenúť, že povolenie na prevádzkovanie snímacích zariadení sa vydáva spravidla na jeden rok a je zo zákona spoplatňované. Do ďalšieho rozporu so zákonom sa môže prevádzkovateľ dronov dostať vtedy, ak ich prevádzkuje v oblastiach s faunou citlivou na hluk. V tom prípade musí mať aj súhlas orgánu ochrany prírody.

3.1.1 Príklady legislatívy pre UAV v iných štatoch

V susednom Česku je situácia podobná ako u nás. Prehľad povinností prevádzkovateľa bezpilotných vzdušných prostriedkov sa dá nájsť na internete¹. Prevádzku bezpilotných systémov upravuje *Doplněk X - Bezpilotní systémy, Předpisu L 2 - Pravidla létání*. Prevádzkovanie UAV do hmotnosti 20 kg na rekreačné a športové účely je možné bez povolenia a evidencie. Pre verejné vystúpenia sa však vyžaduje poistenie. Na komerčné účely sa vyžaduje povolenie a evidencia pre všetky UAV bez ohľadu na maximálnu vzletovú hmotnosť.

V Nemecku je prevádzkovanie UAV právne podchytene dodatkom k existujúcemu Zákonom o leteckých prevádzkach, ktorý definuje UAV ako lietadlá prevádzkované na iné ako rekreačné a športové účely. Nemecký zákon sa podrobne zaobrá aj ochranou dát, ochranou súkromia a ochranou autorských práv. Definiuje aj sankcie v prípade porušenia zákona.²

Ani v Anglicku sa legislatívne obmedzenia netýkajú zariadení, ktoré sa využívajú na rekreačné a športové účely, pokiaľ ich hmotnosť nepresiahne 20 kg. Vo všetkých prípadoch je však nutné dodržiavať zákony na ochranu súkromia a ochranu dát.³

Trochu odlišná situácia je v USA, kde sa UAV využívajú už dlhšiu dobu. Pokročilé technológie autonómnych bezpilotných systémov sa využívajú predovšetkým na vojenské a špionážne účely. Avšak aj využívanie na civilné účely je rozšírenejšie, ako v Európe. Bezpilotné systémy sa využívajú na rôzne komerčné účely, ale patria aj k výbave policajných a záchranných zložiek. Sú známe prípady, kedy drony pomohli pri zatýkaní nebezpečných zločincov.

Prínos dronov a ich potenciál pre rôzne oblasti je nesporný. Vzhľadom na to, že sa množia prípady, kedy aj drony používané na rekreačné a športové účely spôsobujú bezpečnostné riziko pre obyvateľstvo, je otázka prísnejšej legislatívy veľmi aktuálna. Je dôležité, aby sa do tvorby legislatívy zapojili okrem zákonodarcov aj užívateľia a vývojári dronov. Takýto prístup určite prinesie pozitívne výsledky a vytvorí priestor pre užívateľsky priateľné a pre obyvateľov bezpečné využívanie dronov.

4 Etické problémy využívania UAV

Aj keď sa etické problémy v súvislosti s UAV týkajú väčšinou ich vojenského použitia, predsa len existujú otázky, ktorým sa musíme venovať aj pri civilnom používaní dronov.

Rozoznávame niekoľko okruhov problémov, ktoré

¹<http://www.caa.cz/letadla-bez-pilota-na-palube?lang=1>

²<https://www.wbs-law.de/internetrecht/civilian-drones-legal-issues-surrounding-use-50459/>

³<http://www.telegraph.co.uk/technology/2016/04/18/drone-laws-in-the-uk-what-are-the-rules/>

sú podrobnejšie opísané v nasledujúcich podkapitolách.

4.1 Diaľkové a autonómne riadenie pohybu

Predstavme si reálnu situáciu, že diaľkovo riadime dron, ktorý operuje vo veľkej vzdialosti od operátora. Jeho úlohou je zbierať informácie o neznámom prostredí, monitorovať poľnohospodárke alebo lesné plochy, kontrolovať diaľkové vedenia. Takéto činnosti môžu pri súčasnom stave technológií vykonávať drony aj autonómne. Ich prednosťou je, že pri práci nepodliehajú únavе, nerozptyľujú sa okolnosťami a úlohu na ktorú boli určené a naprogramované vykonávajú bezchybne. Dostanú sa do prostredia, ktoré je pre ľudí nebezpečné alebo by vykonávanie práce v taktom prostredí bolo pre ľudí veľmi namáhavé. Zdá sa, že je všetko v poriadku. Aké sú teda etické problémy? Prvý okruh sa týka práce operátora, ktorý je zodpovedný za riadenie UAV vo vzdialom prostredí. Ak napríklad diaľkovo riadený resp. autonómny dron vyhľadáva obete na miestach prírodnej katastrofy a záznam o nájdených obeťach prenáša na operátorské centrum. Ako bude vplývať na psychiku operátora to, že uvidí zblízka utrenie osôb, ktorým nebude môcť reálne pomôcť? Aké etické princípy bude uplatňovať, ak bude musieť voliť ktorým osobám poskytnúť pomoc skôr a ktorým pomoc neposkytnúť vôbec. Záznamy situácie sa budú dať znova prehrať a dodatočne sa bude dať overiť správnosť, resp. ne-správnosť operátorových rozhodnutí, čo bude mať tiež dopady na jeho psychiku.

Druhým okruhom etických problémov v tejto súvislosti je všeobecná dostupnosť, resp. nedostupnosť technológií pre obyvateľov rôznych regiónov. Ich využívanie prinesie prehlbovanie nerovnosti a v konečnom dôsledku aj napäťia medzi obyvateľmi.

Ďalším faktorom je dôvera ľudí k práci autonómnych zariadení. Ak budú na vzdialých územiaci vykonávať činnosť bezpilotné systémy, ich prítomnosť bude miestnymi obyvateľmi prijímaná rozporuplne bez ohľadu na to, aké ciele budú tieto zariadenia plniť. Tu je potrebné aplikovať všeobecne platné etické princípy a nevykonávať činnosti, ktoré by viedli alebo mohli viesť k poškodeniu iných ľudí, fauny, flóry ani majetku.

4.2 Zbieranie údajov

Táto oblasť je pomerne dobre legislatívne rozpracovaná v rámci ochrany dát a ochrany súkromia. Etické problémy, ktoré s tým súvisia sa týkajú rozhodnutí, či a za akých okolností je možné použiť získané údaje pre zaistenie bezpečnosti osôb a vecí. Systémy s umelou inteligenciou zohrávajú nezastupiteľnú úlohu aj v boji proti terorizmu. Vďaka pokročilým technológiám dokážu rozpoznať v dave ľudí tváre, podozrivé aktivity, podozrivé predmety. UAV sa dostanú do nebezpečných priestorov a sprostredkujú nám o nich in-

formácie. Pri plnení úloh však získavajú informácie aj o množstve iných obyvateľov. Aj v týchto prípadoch je potrebné posúdiť, do akej miery sme ochotní vzdať sa súkromia v záujme bezpečnosti.

4.3 Autonómne rozhodovanie

Oblasť autonómneho rozhodovania systémov s umelou inteligenciou patrí k najzávažnejším a najviac diskutovaným otázkam v oblasti roboetiky. Predpokladajme, že dron, resp. autonómny robot všeobecne, je vybavený vyspelým senzorickým systémom a má rozvinuté kognitívne schopnosti. Už dnes je zrejmé že roboty vďaka vyspelým technológiám dokážu omnoho lepšie vyhodnotiť situáciu, ako človek. Sú vybavené zmyslami, ktoré im umožňujú rozpoznať objekty skryté za prekážkou, lepšie sa orientujú v tme, disponujú výpočtovým výkonom, vďaka ktorému vedia veľmi rýchlo a efektívne rozpoznať objekty. Pri výkone mi sie nie sú zaťažené emociami, pokiaľ im ich nenaprogramujeme. A tu je základný kameň problému. Ako naprogramovať rozhodovací mechanizmus robota, aby sa vedel vysporiadať s komplexnými úlohami v zmysle utilitáriánskej alebo deontologickej etiky. Ktorý princíp sa má uprednostniť. Môže robot odmietnuť vykonať príkaz operátora, ak na základe svojich schopností vyhodnotí situáciu inak ako operátor? Problematike autonómneho rozhodovania sa venuje publikácia Vellino a Alaieri (2016). Robot vybavený schopnosťou rozhodovať na základe vlastného zhodnotenia situácie sa považuje za morálneho robota. Sullins (2016) Avšak ako vyvíjať morálneho robota? Existujú rôzne morálne princípy v závislosti od kultúry, sociálneho postavenia, religiozity a geografickej polohy. Vedci experimentujú s rôznymi prístupmi k vývoju morálnych robotov. Podľa Malle a Scheutz (2014), robot musí vedieť rozlíšiť, čo je správne a čo nie je správne. Každé rozhodnutie sa však vzťahuje na konkrétnu situáciu a vyhodnoteenie dobrého a zlého sa môže v jednotlivých situáciach lísiť. Preto sa B. Malle a M. Scheutz pokúšajú zostať akýsi slovník morálnych pojmov⁴. Skupina dobrovoľníkov vyhodnocuje možné aktivity v kontexte nejakej situácie ako správne a nesprávne. Výsledky uložia do sémantickej siete, ktorá ukáže vzťahy medzi jednotlivými aktivitami a kontextom. Takýmto spôsobom vznikne siet, na základe ktorej sa bude dať rozhodnúť, ako veľmi správne alebo nesprávne je vykonanie určitej aktivity v príslušnom kontexte. Ak takúto sieť bude mať k dispozícii morálny robot, môžme predpokladať, že jeho správanie bude odpovedať etickým princípm v danom kontexte. Vytvorenie takéhoto všeobecného slovníka morálnych pojmov je ale nerealistické. Preto sa musíme obmedziť na vytvorenie etických pravidiel pre navrhovanie rozhodovacích algoritmov inteligentných systémov. Pri správnom vytvorení rozhodovacích me-

⁴<http://spectrum.ieee.org/video/robotics/robotics-software/how-to-build-a-moral-robot>, by Kristen Clark

chanizmov je možné predpokladať, že systém s umelou inteligenciou bude rozhodovať nezaujate a uprednostní riešenie, ktoré prinesie osoh viacerým jedincom alebo riešenie, ktorého dôsledky budú priaznivejšie.

4.4 Ochrana životného prostredia

Dôležité otázky ochrany životného prostredia súvisia so znečistením prostredia hlukom a vrakmi poškodených a zničených UAV najmä pri používaní mikro dronov, ktoré budú pracovať vo veľkých skupinách. Takéto drony pri vdýchnutí alebo požití predstavujú vázne zdravotné riziko nie len pre ľudí ale aj pre živočíchy daného prostredia. Týka sa to aj používania robotov pohybujúcich sa vo vodnom prostredí.

5 UAV pre vojenské a obranné účely

Táto problematika má svoje špecifiká a podrobne sa ňou zaoberejú inštitúcie pracujúce vo vojenskej oblasti. Z pohľadu výskumníka v civilnej oblasti je dôležité uvedomiť si, že každý výsledok je zneužiteľný a že nie všetky vojenské aktivity sú nehumánne. Gugliemo Tamburini vo svojom článku Tamburrini (2016) upozorňuje na nebezpečenstvá súvisiace s používaním autonómnych zbraní. Niektoré právne a etické aspekty sú diskutované v článku Kreps a Kaag (2012). V súčasnosti sa na vojenské účely používa široká škála robotov operujúcich v rôznych prostrediach. Ich služby pri strážení hraníc (napri. v Izraeli a v Južnej Kórei) a pri práci v nebezpečných prostrediach, sú pre ľudí mimoriadne prospěšné. Z etického hľadiska sú však veľmi diskutabilné. Predovšetkým je potrebné si uvedomiť, že súčasné technológie sú sice tak vyspelé, že niektorými svojimi schopnosťami predčia človeka, na druhej strane ani najdokonalejšie súčasné stroje nie sú neomylné. To samozrejme nie je ani človek, ale v prípade zlyhania autonómneho systému môžu byť dôsledky omnoho rozsiahlejšie. UAV sa na vojenské účely používajú na získavanie informácií o cudzom území, dopravu rôznych zariadení, informácií a technológií na vzdialené miesta a môžu sa použiť aj na priame vojenské zásahy. V prípade otvoreného vojnového konfliktu sa dá predpokladať, že ich aktivity budú v určitom zmysle humánnejšie. Robot nepodlieha stresu ani zlým emociám. Voči nepriateľom neprejavuje nenávisť, nemá tendenciu páchať zločiny a násilie. Aj v prípade vojenských konfliktov sa UAV dajú použiť na výsostne humánne účely, ako je vyhľadávanie zranených vojakov a ich doprava do bezpečia. Napriek tomu sú oprávnené obavy z používania autonómnych bezpilotných prostriedkov vo vojnových konfliktoch. UAV, ktoré vykonávajú prieskum nad nepriateľským územím môžu obsahovať údaje, ktoré sa pri zachytení nesprávnymi aktérmi stávajú ľahko zneužiteľné. Používanie UAV vytvára nerovnováhu súl a odpornosti-

nych obyvateľov. Aj keď budú roboty vybavené vyspelými technológiami, vždy hrozí riziko, že nesprávne rozpoznajú civilné a vojenské objekty. Pri útokoch na objekty môže dôjsť k ďalším škodám a neplánovaným obetiam na civilnom obyvateľstve. Aký je akceptačný podiel nevinných obetí, aby sme útok vyhodnotili ako úspešný? Podľa medzinárodného dohovoru je zakazané používanie zbraní, ktorých smrtiaca schopnosť je vyššia ako 25%. Napríklad chemické a biologické zbrane sú zakázané, pretože sú príliš efektívne. Môžeme považovať robota, ktorého efektivita sa blíži k 100%, za humánnejšieho len preto, lebo sa nezameriava na civilné objekty? Lin (2011)

6 Záver

Potreba širokého uplatňovania etických princípov v súvislosti s rozvojom robotiky je veľmi naliehavá. Výskum v oblasti roboetiky sa sústredí na krátkodobé aj na dlhodobé ciele. V súčasnosti je potrebné riešiť predovšetkým problémy spoľahlivosti a bezpečnosti robotických systémov, problémy ochrany súkromia ale aj ochranu životného prostredia a zdravia obyvateľstva. Veruggio (2016) Tamburrini (2014) V blízkej budúcnosti je však treba myslieť aj na to, že s rozvojom umelej inteligencie a autonómnych robotov vyvstanú ďalšie etické otázky. Predovšetkým otázky zodpovednosti za konanie a autonómne rozhodovanie systémov s umelou inteligenciou. Rovnako naliehavé sú aj otázky autorských práv a práv systémov s umelou inteligenciou na férové zaobchádzanie. V širších súvislostiach môžeme konštatovať, že technologický pokrok má vplyv na znižovanie prirodzených ľudských schopností (vďaka pamäti našich mobilov strácame schopnosť pamätať si telefónne čísla, znižuje sa vedomosť správne gramaticky písanie a rozpoznávať pravopisné chyby vďaka automatickej kontrole textu). Je dobré, ak si vedci pracujúci v oblastiach umelej inteligencie a robotiky uvedomia svoju zodpovednosť a už pri vývoji inteligentných systémov budú uvažovať o možných dôsledkoch. Napríklad či je správne, aby mali roboty emócie. Ak bude umelý inteligentný systém vybavený určitou škálou pocitov, je alebo nie je správne vystavovať ho zlému zaobchádzaniu zo strany ľudí alebo iných robotov? Kto bude vlastníkom autorských práv pre diela vzniknuté na základe kreatívnych vlastností umelých inteligentných systémov? Ako vyvíjať roboty, aby ich autonómne konanie bolo etické? Ako vyvíjať morálne roboty a aké morálne hodnoty majú rešpektovať? V istom zmysle dokážu systémy s umelou inteligenciou konať etickejšie ako ľudia. Vďaka svojim kognitívnym schopnostiam dokážu lepšie vyhodnotiť situáciu a predvídať dôsledky svojich rozhodnutí v zmysle utilitariánskej a deontologickej etiky. Ich rozhodovacie mechanizmu nebudú začažené emociami a osobnými chybami. Dokážeme však vyvinúť „umelú

morálku“? Nadšenie pre prácu na vývoji stále inteligentnejších umelých systémov nás nesmie zviesť na cestu vývoja systémov, ktoré zmenia našu morálku a etické princípy budované dlhé stáročia. Máme však šancu budovať lepšiu spoločnosť, lepšie porozumieť aj našim slabostiam a využiť potenciál umelých inteligentných systémov v záujme zlepšenia života ľudstva.

Pod'akovanie

Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantovej úlohy č. 2/0154/16 a projektu COST - CA15140, podporeného zo zdrojov MVTS.

Literatúra

DÚ/L001-A/v3 (2015). Rozhodnutie č. 1/2015 z 19.08.2015, ktorým sa určujú podmienky vykonania letu lietadlom spôsobilým lietať bez pilota vo vzdušnom priestore slovenskej republiky.

EASA (2015). Introduction of a regulatory framework for the operation of unmanned aircraft. *European Aviation Safety Agency*.

Kreps, S. a Kaag, J. (2012). The use of unmanned aerial vehicles in contemporary conflict: A legal and ethical analysis. *2012 Northeastern Political Science Association 0032-3497/12 www.palgrave-journals.com/polity/*.

Lin, P. (2011). Drone-ethics briefing: What a leading robot expert told the cia. *Technology - The Atlantic*.

Malle, B. F. a Scheutz, M. (2014). Moral competence in social robots. *IEEE International Symposium on Ethics in Engineering, Science, and Technology. June, Chicago, IL*.

Sullins, J. P. (2016). When a robot is a moral agent. *Ethics in Robotics, International Review of Information Ethics, ISSN 1614-1687, eds Capurro R., Hausmanninger T., Weber K. Weil F.*

Tamburini, G. (2014). On the ethical framing of research programs in robotics. *AI & Soc, DOI 10.1007/s00146-015-0627-2*.

Tamburini, G. (2016). On banning autonomous weapons systems: From deontological to wide consequentialist reasons. *Bhuta N. et al. (eds), Autonomous Weapons Systems: Law, Ethics, Policy, Cambridge University Press 2016*.

Vellino, A. a Alaieri, F. (2016). Ethical decision making in robots: Autonomy, trust and responsibility. *Conference paper*.

Veruggio, G. (2016). Euron roboethics roadmap. *EURON Roboethics Atelier Genoa, 27th of February – 3rd of March, 2006, Scuola di Robotica*.

Wilson, R. L. (2014). Ethical issues with use of drone aircraft. *International Symposium on Ethics in Science, Technology and Engineering, 2014 IEEE*.

Racionálni učitelia a intuitívni manažéri: Interakcia veku a kognitívnej reflexie pri rozhodovaní

Vladimíra Čavojová & Marek Jurkovič

Ústav experimentálnej psychológie - Centrum spoločenských a psychologických vied SAV

Dúbravská cesta 9, 841 04 Bratislava

vladimira.cavojova@savba.sk, jurkovicmarek@gmail.com

Abstrakt

Kognitívna reflexia je schopnosť potlačiť prvú intuitívnu (nesprávnu) odpoveď a zapojiť deliberatívne myšlenie pri produkcií správnej odpovede. Možno ju chápať ako doménovo-všeobecnú kognitívnu schopnosť, ktorá sa môže manifestovať rozdielne v rôznych profesiách. Štúdia porovnáva kognitívnu reflexiu, dispozície myšlenia a rozhodovanie súvisiace s diskontovaním budúcich odmen a riskantnými voľbami u participantov z dvoch profesií - finančnej/manažérskej a učiteľskej. Výsledky ukázali, že učitelia sa líšia od manažérov v preferencii intuitívneho myšlenia a kognitívnej reflexie – učitelia skôrovali vyššie v kognitívnej reflexii a nižšie v preferencii intuícii. Učitelia preferovali v úlohách s časovým diskontovaním trpezlivejšie voľby a v úlohách s rizikom bezpečnejšie voľby. Vek interagoval významne s kognitívou reflexiou len u učiteľov. Starší učitelia s vyššou kognitívou reflexiou vykazovali trpezlivejšie voľby, kým mladší učitelia s vyššou kognitívou reflexiou preferovali okamžité odmeny. Výsledky poukazujú na potrebu preskúmania, ako sa kognitívna reflexia prejavuje v rôznych profesionálnych kontextoch.

1 Kognitívna reflexia a rozhodovanie

V príspevku vychádzame z predpokladu, ktorý formuloval tvorca testu kognitívnej reflexie (CRT) Frederick (2005). Frederick (2005) predpokladal, že kognitívne schopnosti súvisia s preferenciou "trpezlivých" vzdialenejších odmen v porovnaní s okamžitým uspokojením. Test kognitívnej reflexie (CRT) meria podľa jeho autora preferenciu deliberatívneho alebo analytického uvažovania pred intuitívnym heuristickým (a nesprávnym riešením). Medzi inhibíciu okamžitej (a nevyhodnej) odmeny a inhibíciu okamžitého intuitívneho (a nesprávneho) riešenia existuje podobnosť v procese zapojenom v oboch prípadoch - v inhibícii. V našom príspevku sme sa rozhodli replikovať tieto zistenia Fredericka (2005). Nás druhý cieľ bolo replikovať zistenia Fredericka (2005) aj v iných profesijných kontextoch, nakoľko príliš často sa vyskytuje kritika, že väčšina psychologických výskumov sa realizuje na študentoch

vysokých škôl (Henrich, Heine, & Norenzayan, 2010) a ešte k tomu obvykle na študentoch nižších ročníkov psychológie. V rámci tohto príspevku sme sa sice nevyhli tomu, že testovanie bolo vykonané z väčšej časti na študentoch z relatívne bohatej západnej industrializovanej spoločnosti, no pokúsili sme sa zahrnúť aj jednotlivcov z rôznych profesijných oblastí a nejedná sa len o študentov psychológie.

Ďalší cieľ tohto príspevku je totiž aj pilotné preskúmanie rozdielov medzi študentmi rôznych odborov resp. jednotlivcov prajúcich v rôznych profesiách. Kým napríklad v prípade inteligencie (a inteligenčných testov) je známych množstvo poznatkov ohľadom individuálnych rozdielov a ich vplyvu na pracovný výkon alebo ich uplatnenia v rámci napr. výberu pracovníkov alebo vhodnosti povolania podľa výšky inteligenčného kvocientu (Hunter & Hunter, 1984; Schmidt, 1998), v oblasti využitia rôznych kognitívnych klamov v profesionálnej oblasti je však objem literatúry výrazne menší (napr. Corgnet, Gonzalez, & Mateo, 2015; Thoma, White, Panigrahi, Strowger, & Anderson, 2015), hoci ak existujú medzi jednotlivcami rozdiely v podliehaní kognitívnym klamom/heuristikám a tieto rozdiely môžu viesť k rozdielnemu pracovnému výkonu (Corgnet a spol., 2015) alebo rozdielnemu pracovnému správaniu (Thoma a spol., 2015), tak by bolo vhodné pokúsiť sa poznatky v tejto oblasti získať s možnosťou neskôr ich využiť napr. v oblasti výberu povolania (Council, 2015) alebo v iných profesionálnych oblastiach, napr. zistenie vplyvu výkonu profesie na mieru podliehania kognitívnym klamom.

Pre skúmanie kognitívnej reflexie na vzorke učiteľov a manažérov sme mali teoretické aj pragmatické dôvody. Kým manažéri sú relatívne dobre preskúmanou vzorkou aj z hľadiska podliehania rôznym kognitívnym omylem (Ballová Mikušková, n.d., 2017; Krause, 2016), učitelia sú skúmaní oveľa menej, no ich schopnosti analytického myšlenia a nepodliehania prvej, no nesprávnej intuitívnej odpovede, nie sú o nič menej dôležité – skôr naopak. Niektoré výskumy navyše ukazujú, že slovenskí učitelia nemajú príliš vysoké kritické myšlenie (Kosturková, 2013), pričom schopnosť uvažovať dekontextualizovaným spôsobom je jedným z dôležitých

aspektov kritického myslenia. Samozrejme, aj v iných povolaniach je nepochybne dôležitá kognitívna reflexia či kognitívna trpežlivosť (schopnosť počkať si na lepšie výsledky v budúcnosti), preto sme pri výbere vzorky brali do úvahy aj aspekt dostupnosti, čo predstavovalo pragmatický dôvod pre výber týchto dvoch vzoriek profesionálov.

Učitelia aj manažéri môžu mať spoločné oveľa viac, než sa zdá na prvý pohľad. Aby boli úspešní, potrebujú efektívne viest' iných ľudí, ktorí sú voči nim v podriadenom postavení. Trpežlivosť je jedným z najdôležitejších faktorov pre efektívny manažment triedy (Witty, 1947), ale aj korporácií (Ettorre, 1996).

Doterajšie výskumy poskytujú v súvislosti s konceptom kognitívnej reflexie sľubné a pre prax využiteľné výsledky. Miera kognitívnej reflexie napr. pravdepodobne súvisí s mierou prokrastinácie na internete u jednotlivcov z tzv. generácie Y - narodených po r. 1980/90 (Corgnet a spol., 2015). Podarilo sa tiež zistiť rozdielnu mieru kognitívnej reflexie medzi finančnými obchodníkmi a inými povolaniami, pričom obchodníci dosahovali významne vyššie skóre v CRT v porovnaní s jednotlivcami z neobchodných alebo nefinančných povolania (Thoma a spol., 2015).

V rámci nášho pilotného výskumu sme sa rozhodli preskúmať možné kognitívne rozdiely medzi študentmi rôznych odborov (ekonomickej a pedagogického zamerania) a tiež možné rozdiely medzi profesionálmi z pedagogickej a manažérskej oblasti na vzorke zo Slovenska. Konkrétnie sme sa zamerali na oblasť kognitívnej reflexie a tiež oblasť diskontovania odmien, nakoľko bola zistená asociácia medzi týmito dvomi oblastami (Frederick, 2005). Zamerali sme sa tiež na súvislosť s vekom, nakoľko boli zistené isté vekové rozdiely (Campitelli & Labollita, 2010; Frederick, 2005) a na možnú interakciu veku na vzťah medzi diskontovaním odmien a kognitívnu reflexiou. V niekoľkých predchádzajúcich výskumoch sa ukázalo, že racionálita má tendenciu zvyšovať sa s vekom, hoci väčšina výskumov doteraz využívala skôr sebavýpovedové metódy (Ballová Mikušková, Hanák, & Čavojová, 2015; Sladek, Bond, & Phillips, 2010). Ako ďalšia bola u participantov zisťovaná preferencia pre racionálne alebo intuitívne uvažovanie pomocou sebavýpovedového dotazníka.

2 Metódy

2.1 Participanti

Na porovnanie medzi finančnou/manažérskou a učiteľskou doménou sme využili výsledky dvoch predchádzajúcich samostatných výskumov (Čavojová & Hanák, 2014; Čavojová & Jurkovič, 2017). Výskumná vzorka 1 teda pochádzala z výskumu Čavojovej a Hanáka (2014) a tvorilo ju 498 externých študentov Fakulty verejnej správy, ktorí boli rozdelení na 143

participantov v manažérskych pozíciah (ďalej "manažéri") s priemerným vekom 31,21 rokov (M roky praxe = 7,69 rokov; 71,3 % žien) a 355 študentov (ďalej "administratíva") s priemerným vekom 25,3 rokov (M roky praxe = 2,74; 70,9 % žien).

Výskumná vzorka 2 pochádzala z výskumu Čavojovej a Jurkoviča (2017) a tvorilo ju 61 učiteľov s priemerným vekom 44,57 (M roky praxe = 20,10 rokov; 83,6 % žien) a 109 študentov učiteľstva s priemerným vekom 22,04 rokov (M roky praxe = 0; 83,1 % žien).

2.2 Materiál

Participanti vypĺňali slovenskú Racionálno – zážitkového inventára (REI) (Pacini & Epstein, 1999), riešili tri kognitívne úlohy Cognitive Reflective Task a 12 rozehodovacích úloh vytvorených Frederickom (Frederick, 2005). Všetky metódy boli participantom administrované online.

2.2.1 Racionálno – zážitkový inventár (REI)

Dotazník REI (The Rational Experiential Inventory) bol vytvorený na základe Kognitívno-zážitkovej sebateórie: REI meria, ktorý z dvoch paralelných spôsobov spracovania informácií ľudia používajú, racionálny alebo zážitkový. Racionálny systém operuje primárne s vedomou úrovňou a je úmyselný, analytický, verbálny. Zážitkový systém je automatický, nevedomý, holistický, asociatívny, neverbálny a je blízko späť s emociami. Intuícia v REI mapuje skôr schopnosť ako osobnostnú črtu.

Používali sme verziu so 40 položkami (Sirota, 2008), kde 20 výrokov je racionálnych a 20 zážitkových. Celkovo REI mapuje štyri dimenzie: racionálna schopnosť, racionálny postoj, zážitková schopnosť, zážitkový postoj.

2.2.2 Test kognitívnej reflexie (CRT)

CRT (Cognitive Reflection Test) je jednoduchý merací nástroj, ktorý bol navrhnutý Frederickom (2005) tak, aby zachytil špecifické kognitívne schopnosti participantu, konkrétnie hodnoti schopnosť potlačiť spontánnu, intuitívnu, avšak nesprávnu odpoved' (Systém 1) a nahradíť ju uváženou, deliberatívnou a zároveň správnu odpoved'ou (Systém 2).

CRT pozostáva z troch relatívne jednoduchých úloh s možnosťou voľných odpovedí, ktoré majú potenciál evokovať rýchle, intuitívne avšak nesprávne odpovede. Všetky správne odpovede indikujú kognitívny typ (Systém 2), žiadna správna odpoved' indikuje intuitívny typ (Systém 1).

2.2.3 Rozhodovacie úlohy

Použili sme úlohy z Fredericka (2005), ktoré boli zamerané na časové diskontovanie a preferenciu riskantných volieb. Participanti z výskumnej vzorky 2 (učitelia a študenti učiteľstva) vyplnili všetky úlohy, kym participanti z výskumnej vzorky 1 (manažéri a študenti verejnej správy) vyplnili len 12 úloh, v ktorých sa ukázali signifikantné rozdiely vo Frederickovej (2005) štúdii, preto sme na analýzu z oboch vzoriek použili len týchto 12 úloh. Išlo o štyri úlohy na časové diskontovanie, kde mali participanti robiť voľbu medzi okamžitou odmenou a výhodnejšou odmenou niekedy v budúcnosti (napr. Uprednostníte 100 EUR teraz alebo 1100 EUR o 10 rokov?) a 8 úloh na rizikovú voľbu, v ktorých sa participanti rozhodovali medzi istou nižšou výhrou a riskantnejšou možnou vyššou výhrou, no s rizikom straty (napr. Uprednostníte vyhrať 100 EUR naisto alebo 3% šancu vyhrať 7000 EUR?).

3 Výsledky

V rámci výsledkov najskôr uvedieme porovnania všetkých štyroch skupín (učitelia, manažéri, študenti učiteľstva a študenti administratívy) vo výkonových aj seba-výpovedových mierach racionality (CRT a REI) a potom špecificky porovnávame učiteľov a manažérov v rozhodovacích úlohách. Napokon sa zameriame na

analýzu interakcie medzi rokmi praxe a vekom s racionalitou.

3.1.1 Porovnanie skupín podľa veku a zastúpenia pohlaví

Porovnali sme jednotlivé skupiny vo vzorke na základe vekových rozdielov a zastúpenia pohlaví.

Zistili sme, že sa skupiny študentov líšili významne v zastúpení pohlaví, χ^2 (2, N = 610) = 8,10, p = ,004, kym skupiny profesionálov sa významne v zastúpení pohlaví nelíšili, χ^2 (2, N = 221) = 3,56, p = ,059.

Skupiny profesionálov sa tiež významne líšili vekovo ($U = 1709,5$, p < ,001). Rovnako sa tiež skupiny študentov líšili vo veku ($U = 21063,5$, p < ,001). Skupiny učiteľov a pedagogických študentov boli ekvivalentné v zastúpení pohlaví, rovnako aj skupiny manažérov a študentov administratívny (p > 0,05).

3.1.2 Porovnanie učiteľov a manažérov v seba-výpovedové a výkonovej mieri racionality

Ako prvý krok sme porovnávali výkon v REI a CRT medzi manažérmi a učiteľmi, ale aj medzi študentmi učiteľstva a študentmi administratívy. Výsledky sú uvedené v tabuľke 1.

	učitelia		študenti pdg.		manažéri		študenti admin.			
	M	SD	M	SD	M	SD	M	SD	F	p
REI-R	87,82	11,75	73,90	14,46	84,68	13,22	79,27	13,16	21,751	<,001
REI-E	72,00	8,01	72,69	9,46	75,59	14,39	75,65	13,94	2,671	0,046
CRT	1,44	1,09	0,82	1,03	0,99	1,16	0,94	1,15	4,344	0,005

Tab. 1: Porovnanie začiatočníkov a pokročilých učiteľov a manažérov v REI a CRT

Z post-hoc (LSD) porovnaní sa ukázalo, že učitelia aj manažéri sa signifikantne líšia v preferencii racionality od študentov učiteľstva (p < ,001), ale aj od študentov administratívy (p < ,001) a aj študenti učiteľstva sa líšia od študentov administratívy (p < ,001). V preferencii intuície sme nenašli žiadne signifikantné rozdiely ani medzi manažérmi a učiteľmi, ani medzi študentmi oboch odborov.

Vo výkonovom teste CRT mali opäť učitelia najvyššie skóre a signifikantne sa líšili aj od manažérov (p = ,008), aj od študentov oboch odborov (p = ,001). Ostatné tri skupiny (manažéri a obe skupiny študentov) sa v CRT medzi sebou už nelíšili.

3.1.3 Porovnanie učiteľov a manažérov v rozhodovacích úlohách

Z výsledkov uvedených v tabuľke 2 vyplýva, že v dvoch úlohách (a, b) zo štyroch na časové diskontovanie sa učitelia ukazujú byť ako významne viac trpezliví v porovnaní s manažérmi. O 17,4 % viac učiteľov je ochotných počkať mesiac, aby získali o 400 EUR viac, zároveň o 14,3 % viac učiteľov je ochotných počkať aj rok, aby získali o 40 EUR viac. V úlohách, ktoré si vyžadovali čakať na vyššiu odmenu až 10 rokov sa už rozdiely medzi učiteľmi a manažérmi neukázali ako signifikantné.

		učitelia		manažéri		x2	df	p Value	Cramer'sV
Časové preferencie		n	%	n	%				
a	€3400 tento mesiac alebo €3800 budúci mesiac	13	21,3	55	38,7	8,005	3	0,046	0,11
b	€100 teraz alebo €140 budúci rok	42	68,9	119	83,2	8,291	3	0,04	0,111
c	€100 teraz alebo €1100 o 10 rokov	37	60,7	102	72,9	1,00	3	0,153	0,089
d	€9 teraz alebo €100 o 10 rokov	39	63,9	104	72,7	2,192	3	0,534	0,057
<hr/>									
Riskantné voľby									
a	€1000 naisto alebo 90% šanca vyhrat' €5000	33	54,10	69	49,30%	17,661	3	0,001	0,164
b	€100 naisto alebo 90% šanca vyhrat' €500	30	49,20	58	40,60%	8,518	3	0,036	0,113
c	€1000 naisto alebo 75% šanca vyhrat' €4000	48	78,70	105	73,40%	2,054	3	0,561	0,055
d	€100 naisto alebo 75% šanca vyhrat' €200	52	85,20	117	81,8	0,786	3	0,853	0,034
e	€100 naisto alebo 75% šanca vyhrat' €150	58	95,10	131	91,6	0,983	3	0,805	0,038
f	€100 naisto alebo 50% šanca vyhrat' €300	58	95,10	120	83,9	7,628	3	0,054	0,107
g	€500 naisto alebo 15% šanca vyhrat' €1000000	44	72,10	88	61,5	11,209	3	0,011	0,13
h	€100 naisto alebo 3% šanca vyhrat' €7000	55	90,20	129	90,2	5,114	3	0,164	0,087

Tab. 2: Porovnanie učiteľov a manažérov v rozhodovacích úlohách

Naopak, pri úlohách s riskantnou voľbou učitelia v troch úlohách z 8 preferovali istejšiu výhru pred potenciálne vyššou, no rizikovejšou voľbou, a to dokonca aj keď šanca na výšiu výhru bola 90% (úlohy a, b).

3.1.4 Porovnanie pohlaví v CRT a REI subškálach

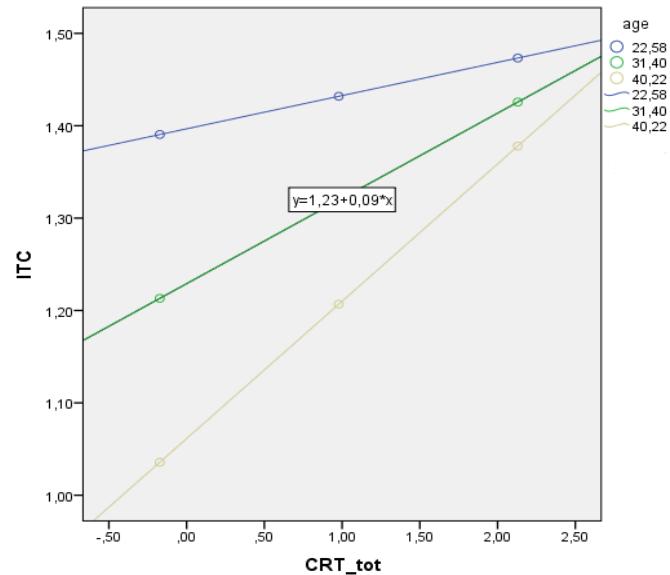
Vzhľadom na nerovnomerné zastúpenie pohlaví vo vzorkách a vyššie zastúpenie žien vo vzorkách s pedagogickým zameraním sme porovnali výsledky CRT a subškály REI aj podľa pohlavia.

Zistili sme významné rozdiely v REI-R skóre ($U = 46797,5$, $p < ,001$), REI-E ($54729,0$, $p = ,007$) aj CRT ($U = 35609,5$, $p = ,001$), pričom muži dosahovali vyššie skóre v REI-R ($M = 84,34$, $SD = 13,12$) ako ženy ($M = 78,81$, $SD = 13,71$) a tiež vyššie skóre v CRT ($M = 1,24$, $SD = 1,12$), než ženy ($M = ,89$, $SD = 1,10$). Ženy naopak dosahovali vyššie skóre v REI-E ($M = 75,81$, $SD = 13,10$) ako muž ($M = 72,59$, $SD = 13,18$).

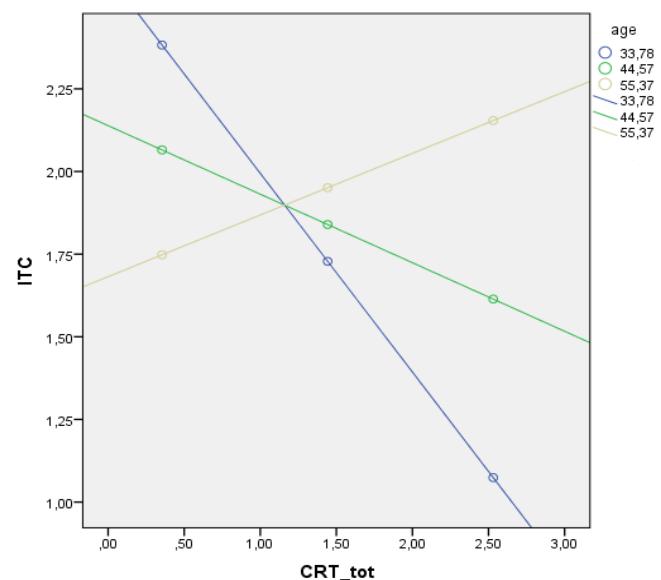
3.1.5 Interakcia medzi vekom a kognitívou reflexiou pri preferencii racionálnejších volieb

Na záver sme testovali interakciu medzi kognitívou reflexiou a vekom a ich efekt na preferenciu trpezlivejších volieb osobitne u manažérov a učiteľov. Analýzu sme robili cez Macro PROCESS v SPSS (Hayes, 2013) a zvolili sme Model 1, kde závislosť premennou bola preferencia trpezlivejších volieb (súčet všetkých trpezlivejších volieb, ITC), kognitívna reflexia bola prediktorem a vek bol moderátorom. Graf na obrázku 1 znázorňuje výsledky interakcie u manažérova graf na obrázku 2 znázorňuje výsledky interakcie u učiteľov.

Analýza ukázala, že vek interaguje signifikantne s kognitívou reflexiou len u učiteľov ($F(3,57) = 2,3100$; $p = ,085$, $R^2 = 0,11$). Starší učitelia s vyššou kognitívou reflexiou vykazovali viac trpezlivejších volieb, kým mladší učitelia s vyššou kognitívou reflexiou vykazovali preferenciu pre okamžité odmeny. U manažérov nebola interakcia medzi vekom a kognitívou reflexiou signifikantná ($F(3,135) = 0,8686$; $p = ,459$, $R^2 = 0,189$).



Obr. 1: Interakcia medzi kognitívou reflexiou a vekom u manažérov



Obr. 2: Interakcia medzi kognitívou reflexiou a vekom u učiteľov

4 Diskusia a záver

Cieľom nášho príspevku bolo replikovať zistenia Fredericka (2005) aj na neštudentskej vzorke z nezápadnej kultúry, no hľavne porovnať efekt kognitívnej reflexie v dvoch špecifických oblastiach profesného výkonu, ako sú manažéri a učitelia.

Na jednej strane, sme potvrdili zistenia iných autorov (Albaity & Rahman, 2014; Čavojová & Hanák, 2014; Frederick, 2005; Hardisty & Weber, 2009; Nofsinger & Varma, 2007; Oechssler, Roider, & Schmitz, 2009), že ľudia skórujúci vyššie v teste

kognitívnej reflexie robia trpezlivejšie voľby a toto zistenie platilo aj u vzorky učiteľov, aj u vzorky manažérov. Na druhej strane, neporovnávali sme len vysoko a nízko kognitívne reflektívnych participantov v jednotlivých voľbách, ale trpezlivejšie voľby sme sčítali a toto skóre sme korelovali so skóre kognitívnej reflexie.

Okrem toho sme tieto zistenia rozšírili aj na dve profesné vzorky – učiteľov a manažérov. V jednotlivých voľbách neboli veľké rozdiely medzi učiteľmi a manažérmi, ale tam, kde sa našli, sa ukázalo, že učitelia sú trpezlivejší a menej ochotní riskovať ako manažéri. Pozoruhodnejší výsledok však bol, že učitelia boli oveľa kognitívne reflektívnejší ako manažéri, hoci v iných výskumoch sa ako reflektívnejší ukázali skôr profesionáli z finančnej oblasti (Akiyama, Hanaki, & Ishikawa, 2014; An, Shi, & Nordvall, 2012; Thoma a spol., 2015). Existuje však niekoľko možných vysvetlení tohto zistenia. V predchádzajúcich výskumu-moch sa napríklad ukázalo, že so zvyšujúcim sa vekom rastie aj kognitívna reflexia a preferencia racionálneho štýlu (Ballová Mikušková a spol., 2015; Čavojová & Jurkovič, 2017; Sladek a spol., 2010). Keďže naša vzorka učiteľov bola v priemere staršia ako vzorka manažérov, rozdiely medzi oboma profesnými skupinami mohli byť spôsobené práve vekom. Tento predpoklad potvrdzujú aj porovnania medzi oboma študentskými vzorkami, kde naopak boli v oboch mierach racionálnejší študenti administratívy.

Skupiny sa tiež líšili v zastúpení pohlaví, pričom skupina študentov pedagogického zamerania obsahovala štatisticky významne vyššie zastúpenie žien. Je možné, že na časti rozdielov medzi skupinami sa podieľali medzipohlavné rozdiely, ktoré boli významné a v zhode s predchádzajúcim výskumom – muži vykazovali vyššie preferenciu pre racionálne analytické uvažovanie (Sladek a spol., 2010), rozdiely boli však vzhľadom na veľkosť štandardných odchýlok minimálne.

Druhé vysvetlenie sa týka skôr homogeneity oboch vzoriek. Kým učitelia predstavovali pomerne homogénnu skupinu skôr skúsených profesionálov, vo vzorke manažérov boli ľudia s rôznymi rokmi praxe, ale zároveň s oveľa rôznorodejšími skúsenosťami, keďže išlo o dospelých externých študentov Fakulty verejnej správy, ktorí si doplnali vzdelanie, aby mohli zastávať rozličné manažérské posty. Na základe našich údajov nebolo možné vzorku manažérov rozčleniť do homogénnejších skupín, príp. sa pozrieť na rozdiely medzi ľuďmi v nižšom, strednom či vyššom manažmente a overiť predpoklad, že na vyšších pozíciah by mali byť ľudia s lepšími schopnosťami kognitívnej reflexie a racionálneho kognitívneho štýlu.

Väčšia veková homogenita vo vzorke manažérov pravdepodobne ovplyvnila aj výsledky interakcie medzi vekom a kognitívou reflexiou, kde sme nenašli efekt interakcie na trpezlivejšie voľby, kým vo vzorke učiteľov interakcia medzi vekom a kognitívou reflexiou zohrávala významnú rolu.

Napriek týmto limitom naša pilotná štúdia, ktorá porovnávala rozdiely v kognitívnej reflexii a rôznych racionálnych voľbách, poukázala na niekoľko rozdielov medzi učiteľmi a manažérmi, ktoré by sa dali zhrnúť do niekoľkých bodov:

1. Najmä starší učitelia vykazujú vyššiu racionality meranú subjektívnymi (REI-R), ako aj objektívnymi (CRT) mierami.
2. Existujú významné rozdiely medzi voľbami učiteľov a manažérov - učitelia sa javia byť ako kognitívne trpezlivejší, kým manažéri sa zdajú byť ochotnejší riskovať, keď možná výhra je výhodnejšia ako istá nižšia voľba.
3. Minimálne u učiteľov dochádzalo k interakcii kognitívnej reflexie s vekom. Kým starší učitelia s vyššou kognitívou reflexiou vykazovali viac trpezlivejších volieb, mladší učitelia s vyššou kognitívou reflexiou vykazovali preferenciu pre okamžité odmeny.
4. Rozdiely boli nielen medzi učiteľmi a manažérmi, ale aj medzi študentmi týchto dvoch oblastí; taktiež sme našli rozdiely medzi nováčikmi (obe študentské vzorky) a pokročilými (obe profesné vzorky) v danej oblasti.

Na záver môžeme konštatovať, že napriek rozsiahlemu výskumu kognitívnej reflexie s použitím testu CRT, sa javí byť dôležité skúmať túto schopnosť v konkrétnych výkonových kontextoch. Rôzne profesie si vyžadujú rozdielnú úroveň kognitívnej reflexie, no zároveň je potrebné rozlísiť, či väčší vplyv na kognitívnu reflexiu má konkrétnie profesné skúsenosti, alebo rastie len s vekom. Táto otázka zostáva zatiaľ otvorená, no môžeme považovať minimálne za povzbudivé zistenie, že je možné ju zvyšovať a cielene podporovať.

Pod'akovanie

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-036 1-12.

Literatúra

- Akiyama, E., Hanaki, N., & Ishikawa, R. (2014). How do experienced traders respond to inflows of inexperienced traders? An experimental analysis. *Journal of Economic Dynamics and Control*, 45(59), 1–18. <http://doi.org/10.1016/j.jedc.2014.05.007>
- Albaity, M., & Rahman, M. (2014). Cognitive reflection test and behavioral biases in Malaysia. *Judgment and Decision Making*, 9(2), 149–151.
- An, N., Shi, X., & Nordvall, A. (2012). *Cognitive Ability and Psychological Biases - Perspectives from Chinese stock individual investors*. Umeå University.
- Ballová Mikušková, E. (n.d.). Rozhodovanie slovenských manažérov so zameraním na využívanie

- intuicie. Bratislava.
- Ballová Mikušková, E. (2017). Intuition in managers' decision-making: qualitative study. *Journal for East European Management Studies*.
- Ballová Mikušková, E., Hanák, R., & Čavojová, V. (2015). Appropriateness of two inventories measuring intuition (The PID and the REI) for Slovak population. *Studia Psychologica*, 57(1), 63–82.
- Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5(3), 182–191.
- Corgnet, B., Gonzalez, R. H., & Mateo, R. (2015). Cognitive reflection and the diligent worker: An experimental study of millennial. *PLoS ONE*, 10(11), 1–13. <http://doi.org/10.1371/journal.pone.0141243>
- Council, N. R. (2015). *Measuring Human Capabilities: An Agenda for Basic Research on the Assessment of Individual and Group Performance Potential for Military Accession*. Retrieved from <http://www.nap.edu/catalog/19017/measuring-human-capabilities-an-agenda-for-basic-research-on-the>
- Čavojová, V., & Hanák, R. (2014). Assessing fluid rationality and its relation to cognitive styles. In T. Ahram, W. Karwowski, & T. Marek (Eds.), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014*. Krakow.
- Čavojová, V., & Jurkovič, M. (2017). Comparison of experienced vs novice teachers in cognitive reflection and rationality. *Studia Psychologica*, 59(3).
- Ettorre, B. (1996). When patience is a corporate virtue. *Management Review*, 28.
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42. <http://doi.org/10.1257/089533005775196732>
- Hardisty, D. J., & Weber, E. U. (2009). Discounting future green: money versus the environment. *Journal of Experimental Psychology. General*, 138(3), 329–340. <http://doi.org/10.1037/a0016433>
- Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A regression-based approach*. New Jersey: The Guilford Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33, 61–83–135. <http://doi.org/10.1017/S0140525X0999152X>
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72–98. <http://doi.org/10.1037/0033-2909.96.1.72>
- Kosturková, M. (2013). Kritické myslenie pedagógov stredných škôl. *Pedagogika.sk*, 4(4), 283–298.
- Krause, R. (2016). Efekt nadmernej sebadôvery v rozhodovacím procese profesionálov. In I. Sarmány-Schuller, L. Pilárik, & I. Gallová (Eds.), *Rozhodovanie v kontexte kognície, osobnosti a emócií V : Zborník príspevkov*. Nitra: Univerzita Konštantína Filozofa.
- Nofsinger, J., & Varma, A. (2007). How analytical is your financial advisor? *Financial Services Review*, 16, 245–260.
- Oechssler, J. J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72(1), 147–152. <http://doi.org/10.1016/j.jebo.2009.04.018>
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76(6), 972–987.
- Schmidt, F. L. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Journal of Personality and Social Psychology*, 74(2), 262–274. <http://doi.org/10.1037/0022-3735.74.2.262>
- Sirota, M. (2008). Individuálne rozdiely v racionálnom usudzovaní. *Kognitívne parametre a komputačné stratégie v „bayesovskom“ usudzovaní v úlohách s odlišnými číselnými reprezentáciami*. (Dizertačná práca). FF UK: Bratislava.
- Sladek, R. M., Bond, M. J., & Phillips, P. A. (2010). Age and gender differences in preferences for rational and experiential thinking. *Personality and Individual Differences*, 49(8), 907–911.
- Thoma, V., White, E., Panigrahi, A., Strowger, V., & Anderson, I. (2015). Good Thinking or Gut Feeling? Cognitive Reflection and Intuition in Traders, Bankers and Financial Non-Experts. *Plos One*, 10(4), e0123202. <http://doi.org/10.1371/journal.pone.0123202>
- Witty, P. (1947). An Analysis of the Personality Traits of the Effective Teacher. *The Journal of Educational Research*, 40(9), 662–671. <http://doi.org/10.1080/00220671.1947.10881565>

Využitie Gärdenforsovej geometrie významu pre uchopenie sémantiky pojmu krásy

Andrej Démuth

Centrum kognitívnych štúdií
Katedra filozofie FF TU v Trnave
andrej.demuth@truni.sk

Abstrakt

Text sa pokúša načrtnúť cestu, ktorou by bolo možné vymedziť a bližšie lingvisticky a sémanticky osvetliť význam a dimenzie pojmu krásy, a to prostredníctvom konceptuálnej analýzy tohto pojmu a jemu blízkych pojmov či dimenzií. Nástrojom na uchopenie významu pojmu krásy pritom bude mapovanie významov pojmov, ktoré ponúka koncepciu sémantických priestorov Petra Gärdenforsa.

1 Úvod

Čo majú spoločné východ Slnka, van Goghova Hviezdná obloha, Bachove kantáty, oči Aishwarye Rai, Michelangelov Dávid, Tádž Mahal či úchvatná scenéria pustej púste v Namíbii? To, čo ich spája, možno opísť ako krásu. Je však obsah týchto objektov natoľko súrody, že ho možno opísť jedným pojmom? Čo je vlastne obsahom a sémantickým významom pojmu krásy?

Mnohí mysliteľia (ako napr. Jerrold Levinson, 2011) sa nazdávajú, že existuje nekonečné množstvo neredukovateľných druhov krásy, a preto aj používanie jedného pojmu na ich označenie je tak trochu metaforické. Prírodná krásy, umelecká, či fyzická prítážlivosť v sebe zahrňajú značne odlišné momenty. Nie lenže označujú veľmi odlišné a svoju povahu len veľmi ľažko súmerateľné objekty, ale aj pocity, ktoré vyvolávajú, sú často charakterizované diametrálnie odlišnými motívmi a rozličným - často až protichodným obsahom nášho prežívania. Možno preto jasne a zrozumiteľne hovoriť o krásach ako čomsi jednotnom, alebo je krás viacero a vo svojej povahе sú nesúmernateľné? Existuje všobec niečo také ako akási jednotná krásy? Alebo je to len abstraktný pojem, ktorým sme sa naučili označovať často veľmi odlišné obsahy a pocity? Je krás všobec vyjadriteľná alebo sú všetky pokusy o jej presné a úplné postihnutie celkom nedosiagnuteľné. Možno nájsť nejaký stály a uchopiteľný obsah pojmu krásy? Aký je vzťah medzi pojмami krásy, nádhera, príjemno, prítážlivosť, vznešeno a ostatnými estetickými kategóriami?

Zdá sa, že v rozmanitých jazykoch predsa len veríme, že existuje niečo, čo označujeme jednotným

pojmom krásy. V dejinách filozofie nájdeme mnoho rozmanitých pokusov definovať krásu. Už samotná existencia viacerých pokusov o jej definovanie svedčí o tom, že jednotlivé definície nepostihujú všetko, čo pod daným pojmom rozumieme. Vzniká dokonca otázka, či je všobec možné krásu zmysluplnie definovať. Pre viacerých filozofov je nemiestne pokúšať sa krásu definovať – akoby to bolo prejavom intelektuálneho nevkusu, pochabým bláznovstvom pokúšajúcim sa o naplnenie sisyfovskej úsilia, či herézou snažiacou sa racionálne uchopiť to, čo sa svojou podstatou rozumu vymyká.

2 Krása a Gärdenforsov model geometrie myslenia

V predkladanom teste sa pokúsim načrtnúť cestu, ktorou, ako dúfam, možno pojem krásy bližšie lingvisticky osvetliť, a to prostredníctvom konceptuálnej analýzy pojmu krásy a jemu blízkych pojmov či dimenzií, s ktorými je spájaný. Pri svojej analýze sa budem opierať o koncepciu sémantických priestorov Petra Gärdenforsa. Gärdenfors (2000, 2004, 2014) predpokladá, že každý pojem možno chápať ako sémantický priestor, ktorý možno študovať a opísť prostriedkami sémantickej geometrie myslenia. Predpokladám pri tom, že pojem krásy predstavuje veľmi komplexný – multidimenzionálny priestor, a preto namiesto pokusu o jeho detailnú a celostnú charakteristiku sa skôr zameriam na analýzu niektorých hlavných dimenzií, ktoré tento priestor vytvárajú a charakterizujú. Jednotlivé dimenzie priestoru predstavujú akési osi rovín – kontinuum, prostredníctvom ktorého možno v nimi tvorennej rovine opísť rozmanité prvky danej množiny (napr. príjemnosť, neutrálnosť až nepríjemnosť; atraktivitu až averziu) a kombináciou týchto rovín aj odlišiť jednotlivé významové a obsahové odtiene, ktoré pri prežívaní pocitu krásy zažívame.

2.1 Subjektivita a objektivita krásy

Asi najzákladnejšou otázkou týkajúcou sa krásy je to, či sa jedná o objektívnu alebo subjektívnu kvalitu

nejakého objektu. Filozofi sa od nepamäti prú o to, či je krása objektívna a od pozorovateľa nezávislá kvalita, alebo naopak, či je jej miesto len v oku (mysli, mozgu) toho, kto ju uzrie či dokáže pozorovať.

Na jednej strane hovoríme o krásе ako o objektívnej vlastnosti vecí, pozorujeme zhodu v estetickom hodnotení objektov, dokonca neraz očakávame akýsi všeobecný súhlas s našim hodnotiacim súdom. Vnímame, že mnogé klasické umelecké diela vyvolávajú zaľúbenie a potechu napriek časom i kultúrami, že existuje móda a všeobecná tendencia vyhľadávania a oceňovania istých objektov.

Na druhej strane však, podstatnou súčasťou krásy je jej subjektívne prežívanie a pocitovanie. Uvedomujeme si, že pri vnímaní krásy vnímame vlastné pocity – zväčša príjemné – ale i to, že tie isté podnetы nemusia u rôznych ľudí vyvolávať pocit záľuby či páčenia sa, a teda, že krása je skôr kvalitou vnemu než objektívnu kvalitou podnetov. Preto je násr vokus často jedinečný, závislý od mnogých, individuálnych – vrodených, ale aj vzdelaním, výchovou či skúsenostou nadobudnutých daností.

Zástancovia filozofického realizmu (napr. Thomas Reid, 1969, p. 755) upriamujú pozornosť na skúmanie parametrov krásnych objektov. Veria, že priamo v ich forme, matérii či v proporciiach tkvie ich výnimočnosť a dokonalosť, ktoré vnímame ako krásne. Preto je potrebné skúmať samotné (krásne) predmety, ich stavbu, štruktúru, ale aj proporcie či kontext. To, čo vnímame ako krásne, je predsa spôsobené štruktúrou a celkovou povahou objektu – podnetu, ktorý afikuje naše receptory (Winckelmann). Napokon, ved' aj z povahy našich estetických súdov vyplýva, že pri krásnych predmetoch predpokladáme krásu ako ich objektívnu vlastnosť, a nielen ako črtu nášho subjektívneho vnímania. Vrávime, že *objekt je*, alebo *nie je* krásny, nielen to, že sa *nám* páči. Predpokladáme, že sa páči aj iným, pretože aj iní vnímajú kvality tohto objektu.

Naproti tomu senzualisti či idealisti chápú, že predmetom našich estetických súdov nie sú veci osobe, ale naše vnemy či predstavy (Hume 2008, 136). Tie sice môžu byť spôsobené kauzálnym (bottom-up) pôsobením objektu, ale vždy zároveň aj stavom a štruktúrou receptorov recipienta, a teda aj individuálnymi odlišnosťami subjektov. Veľká časť antirealistov predpokladá, že objekty vnímania či predstav si sami konštruiujeme, a teda, že predmetom vnímania sú naše predstavy, v ktorých môžeme podstatnú úlohu zohrávať aj predchádzajúca skúsenosť, učenie, či očakávania (top-down procesy). Preto je celkom pochopiteľné, že pri pohľade na ten istý objekt nemusíme disponovať (a často ani nedisponujeme) tými istými vñemami a predstavami. Násr vokus je do značnej mieru subjektívny nakoľko zohľadňuje individuálne odlišnosti recipientov a predmetov ich vnímania. Je teda vobec možná jednotná veda o krásnom, alebo je vokus rýdzo individuálnej a subjektívnej záležitosťou?

Nazdávam sa, že väčšina súčasných mysliteľov predpokladá kompromisné vplyvy na vznik estetického zážitku. „Vnímať, že niečo je krásne a zároveň tvrdiť, že je to nepohnuté a nedotknuté skúsenosťou, sa zdá byť protirečivé“ (McMahon, 2001, p. 228). Vnímať niečo ako objekt a nepredpokladať pritom vplyv subjektu je problematické. O to viac, že pocit krásy nie je len vyjadrením nejakej kvality objektu, ale vždy zároveň aj reflexiou vlastného stavu subjektu, ktorý objekt vníma – je reflexiou nášho vzťahu (pocitu) k (z) objektu – reflexiou toho, či sa nám objekt (predstava) páči. Pocit zaľúbenia, príjemnosti či slasti je subjektívou kvalitou. Kant preto uvažuje o krásе prostredníctvom reflexívnej súdnosti – schopnosti reflektovať pocit potešenia, ktorý vnem či predstava v subjekte vyvoláva.

Predmetom estetického súdu je teda vnem alebo predstava objektu. Prítom je takmer irelevantné, ako tento vnem či predstava vznikajú. Môže ísť o aktuálny vnem napr. obrazu, alebo spomienku naň, či voľnú tvorbu fantázie, ktorá objekt pred našu myseľ pred/postaví (vorstellen). Predmetom estetického súdu preto môžu byť aj objekty, ktoré fyzicky reálne neexistujú. Existujú však v mysli a sú myslou nazenané. To, čo posudzujeme ako krásne, nie je samotný predmet či dokonca jeho existencia, ale skôr štruktúra tohto predmetu v predstave, resp. jej schopnosť, alebo neschopnosť vyvolávať príjemnosť a pocit záľuby.

To, že predmetom estetickej skúsenosti je vnem či predstava (teda mentálne objekty), však ešte neznamená jeho nereálnosť či rýdu subjektivitu. V zmysle Kantovej *Kritiky čistého rozumu*, väčšina objektov, s ktorými sa stretáme, je empiricky reálnych, avšak zároveň aj transcendentálne ideálnych. Takou, okrem času a priestoru, je napríklad farba, chut', ale aj krása. To znamená, že pocit záľuby, ktorý sprevádza krásne objekty v našom vedomí, je empiricky celkom skutočný a má aj objektívnu štruktúru. Odhliadnuc od nášho vedomia je však pocit krásy nezmysluplný. Krása je niečo, čo je krásne pre niekoho a sama osebe (odhliadnuc od akéhokoľvek subjektu) nedáva zmysel. Zároveň však existuje v rovine vedomia a ovplyvňuje naše vnímanie a konanie. V tomto zmysle je empiricky reálna. Ako možno teda robiť vedu o niečom, čo je predmetom „len“ subjektívnej predstavivosti a pocitov? Jednou z možných cest môže byť stratégia, ktorou sa vydávali už pytagorejci.

Pytagorejci verili, že podstatou krásy je harmonickosť a proporcionalnosť objektu, resp. vzťahy medzi jednotlivými prvkami a spôsob ich organizácie (Livio, 2007, p. 144–165). Vzťahy však možno skúmať ako na samotných materiálnych objektoch, tak aj na ich predstavách. Krása (podľa nich) môže byť poňímaná ako relačná vlastnosť, a tú (podľa pytagorejcov) možno (podobne ako harmóniu) vyjadriť geometricky či dokonca číslami.

V prípade fyzických objektov pomeriavanie predmetov a vyjadrovanie vzťahov môže byť celkom

jednoduché, čo dokumentovali aj renesanční umelci (Vitruvius, Durer). Tí za predmety považovali materiálne objekty a ich fyzické parametre porovnávali. Ak teda chceme vedieť, ktoré objekty, alebo ich prvky sú pre vnímanie krásy podstatné, možno sa zamerať na analýzu parametrov a štruktúry objektov, alebo aspoň (napr. eyetrackerom) skúmať to, kam sa dívame a ktoré prvky sú klúčové, keď vnímame niečo, ako pekné. Táto cesta je vlastne cestou realizmu – predpokladajúcou, že predmet estetického súdu možno viac-menej stotožniť s fyzikálnym objektom.

Idealisti – ako Baumgarten či Kant – verili, že predmetom estetického súdu nie je vec, ale predstava formy objektu. Ako ale možno skúmať štruktúru a parametre predstavy? Jednou z možných ciest by bolo predstavy zhmatniť, alebo vizualizovať. Takéto objekty je potom možné merať a porovnávať, podobne ako vyšie spomínané materiálne objekty. Druhou, opačne zameranou cestou, je metóda, zameriavajúca na opis štruktúry vnímanej skúsenosti – predstavy – a jej povahy, prostredníctvom merania a hodnotenia odlišných reakcií probandov na rôzne podnety, alebo ich predstavy v závislosti od objektívne preukázaťelných odlišností fyzikálnych podnetov (napr. iné parametre fotografií), respektívne od odlišností našich reakcií a hodnotení jednotlivých predstáv. Predmetom takejto subjektivisticky orientovanej metódy je preto preferencia charakteristiky obsahu zážitku (aké to je?) pred objektivistickým opisom štruktúry objektu, ktorý skúsenosť vyvoláva (čo to je?). V takomto prístupe sa možno zamerať na fenomenologický opis estetickej skúsenosti – na to, čo prežívame, keď sa nám niečo páči, ako to prežívame, ako sa to líši od iných pocitov, ktoré mávame. takomto prístupe (z prvej osoby) sa potom možno zameriavať aj na problémy, čo pre nás krásu znamená a čo je charakteristickou črtou estetickej skúsenosti.

Predkladaný model sa pokúša o skĺbenie oboch prístupov. Základným momentom dimenzie objektivity a subjektivity krásy je, či recipient väčšmi pocíťuje ako predmet svojej skúsenosti vlastný pocit a pocítovanie (reflexivita záľuby a páčenia sa), alebo je objektom estetickej skúsenosti celkom pohltencí a neuvedomuje si subjekt a vlastné pocity ako pocity, ale vníma niečo, čo považuje za kvality objektu. Miera subjektivity a objektivity krásy pritom môže objekt od objektu kolísat.¹ Zdá sa, že pre niektoré percepcie späť s pôsobením podnetov na vonkajšie receptory (napr. zrakové vnímanie) máme častejšie tendenciu krásu objektivizovať a pre intelektuálne či vnútorné orientované podnety (pocity telesnej slasti) máme zasa tendencie krásu subjektivizovať. Vzťah medzi subjektivitou a objektivitou estetickej skúsenosti možno pritom vnímať ako isté kontinuum, v ktorom sa možno plynule pohybovať.

¹ To, že kolíše od subjektu k subjektu, dokazujú aj uvedené príklady z dejín estetiky.

2.2 Príjemnosť – nepríjemnosť estetickej skúsenosti

Druhou dôležitou charakteristikou krásy je jej kvalitativna dimenzia. Kant bol presvedčený, že estetický súd je primárne späť s reflexiou toho, či daná predstava v nás vyvoláva, alebo nevyvoláva pocit záľuby. (Kant, 1975, §9). Pocit krásy je pre neho reflexívnym pocitom, alebo aspoň pocitom, ktorý sa uvedomovaným môže stať. Je otázne, či možno nazerať krásny objekt, ktorý sa nám páči a nevedieť, že sa nám páči.² Zdá sa preto, že vnímanie krásy je vždy (viac-či menej) späť s vnímaním vlastných subjektívnych stavov, osobitne toho, či daný objekt – predstava – v nás vyvoláva príjemné, alebo nepríjemné pocity. Príjemnosť, resp. nepríjemnosť, sú pre mnohých klúčovou dimensiou estetickej skúsenosti.

Zdá sa, že pocit krásy je vo všeobecnosti spájaný s pozitívnymi pocitmi. Intenzita prežívanej príjemnosti, alebo jej stupňov, môže byť však veľmi rozmanitá. Niekoľko je krásu prežívana ako veľmi intenzívny pozitívny pocit. Vtedy hovoríme o rozkoši, slasti či pôžitku. Inokedy je to skôr pozitívny a prijemný pocit, ktorý hodnotíme kladne, ale uvedomujeme si, že jeho intenzita nie je maximálna. Napokon samotný slovenský termín „príjemný“ vyjadruje práve to, že sa jedná o „jemný“ pocit, a nie maximálne intenzívny pocit.

Príjemné pocity sú relatívne mierne a nevysznačujú sa privékou intenzitou. Nevysznačujú sa ani extrémnou štruktúrou v zmysle prítomnosti veľkých kontrastov, ruptúr a nevšedných charakteristík. To, čo ich spája, je, že nie sú rušivé, ale na druhej strane ich predsa len vnímame. Neutrálny pocit je charakteristický tým, že si ho zväčša neuvedomujeme, respektívne, že si neuvedomujeme žiadnu jeho charakteristickú črtu. Pri príjemných pocitoch si však uvedomujeme ich „príjemnosť“ a teda to, že sa nám páčia, a že nám robia dobre. Pri vnímaní krásy si zväčša uvedomuje práve to, že predmet sa nám páči a jeho predstava nám robí dobre.

Príjemné však nie je príjemným vždy a za každých okolností. Fyziológia nášho vnímania spôsobuje, že vnímanie pôsobenie toho istého podnetu s tou istou intenzitou sa s meniacou dĺžkou jeho pôsobenia tiež mení. A tak opakovana expozičia tých istých podnetov, ktoré spočiatku vnímame ako príjemné, môže v dôsledku častého opakovania svoju príjemnosť postupne stratíť (habituačiou znecitlivieme), alebo môže spôsobiť úplnú zmenu subjektívneho

² Pre mnohých kantovcov nemožno pocíťovať krásu a nevedieť, že ju pocíťujeme, rovnako ako nemožno mať bolesť a nepociťovať ju. K podstate bolesti patrí, že je pocítovaná. To, čo ju spôsobuje, nie je bolesť, podobne ako svetlo dopadajúce na sietnicu nie je farba. Krásu je preto kvalita vnemu a hoci si spočiatku nemusíme uvedomovať, že na nás niečo pôsobí príjemne, o pocite krásy možno hovoriť až vtedy, keď si ho uvedomujeme.

hodnotenia príjemnosti. Pôvodný podnet sa tak stáva nepríjemným, respektíve podnet či predstava vyvolá nepríjemnú až averzívnu reakciu. To sa stáva, ak sme vystavení pôsobeniu nejakého podnetu pridľho, či príliš často – zovšednie, alebo sa nám až znechutí. K príjemnosti teda patrí správna miera – a to ako čo do intenzity, tak aj čo do trvácnosti, či frekvencie. Podobne aj s krásou. To, čo vnímame ako krásne takým po dlhodobej expozícii už byť nemusí. Zgýčovie. Stane sa neutrálnym alebo až nepríjemným.

Dimenziu príjemnosť – nepríjemnosť možno teda taktiež vnímať ako kontinuum a niektoré predmety či predstavy v nás môžu vyvolávať čo do intenzity rôzne príjemné pocity od tých najextrémnejšie pozitívnych, cez mierne príjemné, neutrálne, ba dokonca až po tie najnepríjemnejšie. Tie sú akoby späté s tým, čo považujeme za opak krásy – škaredosť a ošklivosť. Od škaredosti očakávame, že sa nepáči, že vyvoláva nepríjemné pocity, averziu a pocity, ktoré nechceme.³ Je prítom logické predpokladať, že to, čo vnímame ako krásne, nemôžeme zároveň v tom istom čase pocitovať ako škaredé et vice versa, čo dokumentujú aj viaceré neuroestetické výskumy.⁴

Mohlo by sa zdať, že príjemnosť – nepríjemnosť je kardinálnou charakteristikou krásy. Od toho, čo je krásne, očakávame pozitívne pocity, od toho čo je škaredé – nepríjemné pocity. Akoby bolo nemožné vnímať niečo ako krásne a mať pri tom nepríjemné pocity, či naopak, vnímať ošklivosť a pocitovať slast'. Väčšina filozofov preto verí v kauzálnu späťnosť potešenia a krásy a otázkou pre nich je len diferenciácia esteticky príjemných pocitov od iných druhov (neesetickej) slasti. Navyše, zástancovia tohto princípu slasti vyžadujú, aby potešenie, ktoré krásny objekt v nás vyvoláva, nebolo spôsobené nejakými osobnými, materiálnymi či inými konkrétnymi a zainteresovanými potrebami, ale naopak, podstata onej príjemnosti estetického pocitu musí spočívať v samotnej forme a štruktúre objektu – predstavy a v nezaujatom a nezainteresovanom nazeraní objektu. Hoci z evolučného hľadiska dôvody estetického potešenia a iných foriem slasti môžu a pravdepodobne vychádzajú z tých istých koreňov.⁵

Zdá sa teda, že vo všeobecnosti medzi estétmi prevláda spájanie krásy so subjektívne príjemnými pocitmi, hoci na druhej strane, existujú aj názory, že

³ O istej mentálnej atraktívite škaredosti pozri ďalej.

⁴ Semir Zeki a spol. preukázal(i) existenciu neuronálnych centier aktivovaných pri vnímaní krásy a iných pri vnímaní ošklivosti. Obe centrá sú súčasťou toho istého systému odplát a Zekho výskumy (2004, 2013) predpokladajú bipolárne (inhibitórne) pôsobenie centier pre krásu a škaredosť v mOFC.

⁵ Nazdávam sa, že evolučným dôvodom slasti (akejkoľvek), je suportívne pôsobenie na vyhľadávanie a preferovanie podnetov a spôsobov správania vedúcich k žiadúcemu cieľu – zvýšeniu pravdepodobnosti prežitia jedinca a jeho potomstva.

vraj možno vnímať krásne objekty a nepociťovať pri tom žiadne potešenie, nevraviač o tom, že možno pocitovať potešenie a nevnímať niečo ako krásne. Na problematickú oddeliteľnosť krásy s inými charakteristikami ako je len „príjemnosť“ upozorňuje aj bežný jazyk, keď v hovorovej reči neraz zamieňame krásu s napr. s dobrom. Hovoríme o tom, že, ak je niečo dobre namaľované (formálne či remeselne), je to „krásne“ namaľované. A to aj vtedy, ak je tým „krásne namaľovaný“ niečo samo o sebe škaredé. Podobne krásna vec môže byť nepekné (zle) namaľovaná, a teda krásu sa neraz spája aj s inou dimensiou a tou je dokonalosť a exkluzivita.

2.3 Dokonalosť – nedokonalosť

Dôvody zamieňania krásy s dobrom spočívajú v antických (platónskych a epikurejovských) koncepciach. Platón bol presvedčený, že krásny objekt je ten, ktorý najväčšmi participuje na idei, ktorú znázorňuje. Každá idea teda môže byť znázornená rozmanitými spôsobmi, ale len niektoré ju vystihujú dobre. Ak je daný spôsob znázornenia či zhmotnenia idey výstížný, hovoríme, že je to dobre vyjadrené, resp. napr. krásne zahrané či pekne namaľované, povedané. Vyjadriť teda niečo krásne, znamená vyjadriť to dobre, obsahovo aj formálne vystihnúť všetky podstatné črtky – urobiť to remeselne zručne. Krásne, napríklad, možno zahrať skladby, ktoré samy osebe nemusia byť ešte krásne. Krásne možno namaľovať izbu, ktorá nie je ani krásna, ani nepekná, ale dokonca, ako píše Eco, krásne možno namaľovať aj ošklivosť či diabla. Krásne namaľovať ošklivosť neznamená dodat' jej punc krásy,⁶ práve naopak. Krásne je vyjadrená vtedy, keď dokonale vystihuje samotnú ošklivosť, keď jej nič neuberá, ani neskrýva, keď ju vyjadruje takú aká je, živo a farbisto, so všetkými jej nuansami a odtieňmi – prosté keď ju dokonale vystihuje.

Dokonalosť vyjadrenia sa netýka len remeselného zvládnutia, ale súvisí skôr s formou. Krásne môžu byť aj obyčajné veci – rastliny, predmety, ktoré nás obkolesujú, ale ktorých krásu si bežne neuviedomuje, pretože je „zakrytá“ nánosom nevhodného kontextu, matérie či inými determinantmi. Na to, aby sme

⁶ V tejto súvislosti je pozoruhodnou otázkou práve skrášľovanie a zdokonaľovanie vecí. F. Schiller v úvahách *O pôvabe a dôstojnosti* píše, že pôvab je vôlevá krásu, ktorú možno aj architektonicky nedokonalým veciam pridať, aby sa stali krásnejšími. Pôvabné, napr. pohyby, sa takými stávajú až potom, čo si vôlevé, rozumové pravidlá, ktorými sa riadi náš pohyb, osvojíme až do takej miery, že sa stanú našou druhou prirodzenosťou a my dané pohyby vykonávame akoby spontánne. Dodať punc krásy teda znamená, prekryť nedostatky niečim, čo hodnotíme ako krásne, a to v takej miere, že to krásne prevyšuje ošklivosť, či nedokonalosť. V tomto zmysle je skrášľovanie zakrývaním skutočnosti, respektíve jej prekrývaním novým – krásnym nánosom. Je teda zdokonaľovaním nedokonalej veci.

uzreli ich krásu (Krása je v oku toho, kto ju uzrie), je potrebné zbaviť ich týchto nánosov – odkryť ich a uzrieť ich v ich formálnej nahote. Takto si možno uvedomiť napríklad dokonalý tvar šálky či iného úžitkového predmetu, rastliny či línií a proporcii v prírode.

Dokonalosť však predpokladá poznanie účelu (dokonalý na čo?). V platónskom ponímaní ide o poznanie idey, v kognitivistickom o poznanie účelu, pre ktorý je daná vec vyformovaná tak, ako je, aby plnila či dosahovala svoj účel. Biológovia sa neraz nadchýňajú účelnosťou tvarov a štruktúr v prírode, ak pochopia mechanizmy, ktorými príroda dosahuje svoje ciele. Umelecký kritik zas nachádza potešenie tam, kde identifikuje formu účelnosti, ktorou umelec dosiahol účel svojho snaženia, a sice záujem a pocit páčenia sa umeleckého diela. Účelom umeleckých diel je (podľa Kanta) práve páčenie sa. To, či a ako umelci tento ciel dosahujú, spočíva práve v rozličnej forme účelnosti, ktorou sú ich diela koncipované a tým, do akej miery je zvolená forma účelnosti vhodná či účinná pre dosahovanie daného ciela. Preto je práve forma účelnosti klúčovým pojmom Kantovej estetickej súdnosti. Rovnako je však dôležitým pojmom pojem dokonalosti.

Dokonalosť znamená, že daný ciel nemožno dosiahnuť lepším spôsobom. Dokonalé diela teda predstavujú najlepšie možné vyjadrenie danej predstavy či idey. Ak sú dokonalé, nemožno na nich už nič vylepšiť, pridať či uberať. Akýkoľvek ďalší zásah ich kazi. Perfectio je teda vyjadrením úplnosti, je vyjadrením naplnenia danej idey v maximálnej možnej mieri, respektívne vyjadrením toho, že to, čo vnímaleme, alebo si predstavujeme, je v takej podobe, že si nevieme predstaviť, žeby to mohlo byť ešte lepšie, krajsie. Pojem dokonalosti je výrazom akéhosi vnútorného pomeriavania – porovnávania vnímanej predlohy s akým ideálom či predstavou, ktorá reprezentuje najkrajší objekt daného druhu, aký si môže subjekt predstaviť. Porovnávanie a prepočítavanie je však charakteristickým znakom aj inej dimenzie krásy.

2.4 Miera výskytu: Exkluzivita a všednosť

Z povahy úplnosti a dokonalosti vyplýva, že najlepšou zo všetkých možných predstáv či predlôh môže byť len jedna. Častým aspektom úvah o kráse (a najmä o kráse v umení) je preto otázka výnimočnosti a všednosti. Zdá sa, že všedné veci, napriek tomu, že sú dobre urobené a vyvolávajú pozitívne pociaty, zväčša nedokážu produkovať pocit intenzívnej záľuby. Naopak, ak aj tento pocit vyvolávali, postupne ho vyvolávať prestávajú. Spočíva to v tom, že pod vplyvom ich častého či dlhodobého pôsobenia si na ne postupne zvykáme a oni tak strácajú čaro dráždenia. Jednoducho sa ich presýtme. Z rýdzo fyziologickej hľadiska je to preto, že akákoľvek dlhodobá aktivita vyvoláva únavu a aj stimulácia našich orgánov a či saturácia túžob vedie k ich znecitlivovaniu.

David Hume veril (Hume, 2008, p. 144), že podstatou vkusu je vzájomné porovnávanie skúseností. Každý subjekt sa počas života stretne s rozmanitým počtom a druhom skúseností. Ak ich potom následne porovnáva, vytvára si akési neustále sa dopĺňajúce kontinuum zážitkov a vnemov, z ktorých niektoré považuje za príjemné a dobré, iné zas za menej príťažlivé či žiaduce. Toto kontinuum je základom pre tvorbu idey normy, ktorú si možno predstaviť ako hľadanie priemeru, či váženého priemeru kvality jednotlivých zážitkov. V estetickej oblasti je to hľadanie toho, čo je krásne v porovnaní s tým, čo za také už (alebo ešte) nepociťujeme. Je samozrejmé, že s postupom času a s narastajúcim počtom nových skúseností sa mení a posúva aj naša norma estetického vkusu.

Humova predstava vkusovej normy teda predpokladá, že za krásne považujeme len tie veci, ktoré vyvolávajú skúsenosti, ktoré sa nachádzajú nad hranicou všednosti (pozitívne deviácie). Krásne veci nie sú bežné a obvyklé – naopak, sú to práve tie veci, ktoré sa zo štandardu vymykajú. Svojou povahou sú excencké.⁷

Thomas Reid preto uvažuje o excellentnosti ako o jednom z klúčových kritérií krásy. Excellentnosť charakterizuje krásu nielen preto, že krásny objekt patrí k tým najlepšie zrealizovaným a najsilnejšie spôsobilým vyvolávať v nás príjemné pocity, ale aj preto, že najkrajšie veci vyvolávajú najsilnejšie estetické zážitky (Reid, 1969, p. 760).

Jednou z charakteristických črt krásy je jej relatívnosť. Rozmanité predmety vnímame ako krásne vždy v nejakom kontexte. Ak sa objaví niečo, čo sme dosiaľ vnímali ako krásne, v prítomnosti ešte krajsieho objektu, pôvodný predmet ako tak krásny už nevnímame. Je to tým, že v porovnaní s krajsím objektom svojou krásou už tak nevyniká. To je možno dôvod, prečo nám mnohé objekty postupne zovšednejú – už nevynikajú v rámci množstva navnímaných skúseností, alebo v porovnaní s ďalšími – novými. Pritom nemusí ísť len o reálne zažité skúsenosti, ale aj o to, čo si dokážeme len predstaviť, napríklad tvorivosťou fantázie, alebo na základe opisu skúseností iných jedincov, napríklad expertov, ktorí rozširujú diapazón našich predstáv o ich vlastné skúsenosti (Démuth, 2016).

Problém excellentnosti možno vysvetľuje skutočnosť, prečo množstvo objektov možno vnímať ako krásne (napríklad tvar listu, kvetu a pod.), ale v porovnaní s inými exemplármi, napríklad krásneho umenia, si ich krásu neuvedomujeme. Opakom relatívnosti krásy je absolútна krása, ktorá si svoj punc výnimočnosti a dokonalosti ponecháva nedotknutý nezávisle od akéhokoľvek iného predmetu a hodnotenia. Absolútna krása predpokladá, že jej

⁷ O štatistikom mechanizme tvorby vkusového súdu pozri: Démuth (2016).

schopnosť vzbudzovať páčenie sa je nezávislá od kontextu, doby a iných okolností; že je nadčasová. Klasické umelecké diela si neraz nárokuju na punc takejto nadčasovosti a nadkultúrnosti.

Zaujímavým aspektom oboch dimenzií (dokonalosť ako plnosť a exluzivita ako výnimočnosť) je, že sú principiálne založené na porovnávaní. Natíska sa preto otázka, či je krása relačnou vlastnosťou vecí, ktorú možno vyjadriť adjektívom (pekný, krásny, najkrajší), alebo je niečím substancionálnym (krásu ako podstatné meno), charakterizujúcim objekt ako taký, nezávisle od iných. Kým v prípade príjemnosti možno očakávať, že príjemné vnímame iba voči možnému opozitu (nepríjemné) s možnosťou stupňovania oboma smermi (gausovské rozloženie pocitov), dimenzie dokonalosti a exluzivity akoby opisovali hyperbolickú, resp. lineárnu postupnosť. Ich základom je (až geometricky) početnejší výskyt nedokonalých vecí smerujúcich k narastajúcej mieri dokonalosti až k jedinej dokonalej (ak je vôbec dosiahnutel'ná), resp. štatistické rozloženie výskytu javu od najpočetnejšieho k zriedkavému až ojedinelému (ktorý existuje). V prípade príjemnosti a nepríjemnosti pritom neuvažujeme ako o stupnici toho istého (nepríjemnosť nevnímam ako absenciu príjemného, ale ako úplne iný, opačný pocit s odlišnými charakteristikami), v prípade dokonalosti a nedokonalosti máme dočinenie skôr s prítomnosťou a neprítomnosťou tej istej charakteristiky (nedokonalosť ako absencia dokonalosti) a v prípade exkluzivity len vyjadrenie ojedinelosti výskytu toho istého javu voči všednosti.

Obe dimenzie (dokonalosť aj exkluzivita) sú preto principiálne kontinuom, v ktorom sa možno pohybovať od minimálneho stupňa k maximu et vice versa.

2.5 Simplicita a komplexnosť

Mnohí antickí i stredovekí autori venovali pozornosť otázke komplexnosti a simplicity krásy. Platonici verili, že krásne objekty sú krásne nielen preto, že sú symetrické a pomerné, ale najmä preto, že ich dokáže-me uchopiť v ich celistvosti ako harmonické. Preto uvažovali o tom, či ku krásce patrí jednoduchosť, alebo skôr naopak, či preferujeme zložité útvary. V antických dobách nájdeme zreteľné preferencie kruhu pred elipsou či inými neurčitými krvkami, kruhu či kocky pred mnohouholníkmi a mnohostenmi, rovnostranného trojuholníka pred rovnoramenným, a pod. Krásu sa pre Grékov snúbila so symetriou a relatívnu jednoduchosťou. Zložité a neurčité útvary boli spájané s nedokonalosťou či dokonca s opakom krásy.

Možno to súvisí s tým, že neurčité geometrické či iné útvary sa nedajú celkom dobre uchopiť a následne replikovať. V tejto súvislosti možno krásu chápať intelektuálne – ako nahliadnutie *ordó* – systému, v ktorom sú jednotlivé prvky, či ich časti navzájom organizované. Pokiaľ je systém jednoduchý a preukázateľný, pocitujeme intelektuálne uspokojenie

z jeho objavenia a ľahko porovnávame, či sa všetky prvky naozaj hodia do celku. V tom vari spočíva podstata pytagorejského ponímania krásy ako primeranosti a proporčnosti, ale i shaftesburyovské *fitness* – primeranosť a vhodnosť daného prvku k ostatným. Na to, aby boli veci primerané, však potrebujeme poznáť celok a algoritmus jednotlivých pomerov, ako napríklad pravidlo zlatého rezu, či Polykteinov *Kánon* proporcií jednotlivých častí tela, ako uvádzajúce Vitruvius.

Primárnosť celku pred časťami bola častou tému stredovekých úvah. Tie upozornili na spor medzi čírou jednoduchosťou a harmóniou odlišných, často kontrastných častí. Prostá jednoduchosť pôsobí nerušivo a jej algoritmus nie je problém odhaliť. Na druhej strane však neraz pôsobí nudne a nezaujímavo. Naproti tomu, členitosť a kontrastnosť jednotlivých častí, ich komplementarita a vzájomná harmonickosť pútajú pozornosť a zamestnávajú myseľ tak, že pokiaľ v ich vzťahu dokážeme nájsť systém, ktorý odhaluje ich vzájomné dopĺňanie sa a pôsobenie, tak, že celok pôsobí napriek zloženiu z častí harmonicky a vyvážene, zväčša sa dostavuje záujem a pocit príjemného uspokojenia. Miera komplexnosti však musí byť primeraná, pretože veľmi zložité kompozície môžu viesť oko a myseľ k neschopnosti uchopiť jednotlivé časti a nachádzať medzi nimi relevantné vzťahy, čo pocit krásy oslabuje, či dokonca priamo znemožňuje. Príkladom takéhoto veľmi komplexného a členitého ponájatia krásy bolo barokové umenie. Objavenie nekončna (matematického i reálneho) neraz viedlo k naznačeniu nekonečných postupností a úrovni uchopovania jednotlivých objektov (v hudbe aj v maliarstve napr. prostredníctvom repetícií a nekonečných slučiek, ornamentov a pod.) a napriek snahe o symetriu neraz neumožňovalo percipientom uchopiť všetky prvky a docieliť pocit krásy, ktorý bol neraz nahradený dojmom veľkoleposti a vznešenosťi. Podobne pri vrcholnej arabskej architektúre môžeme zostať zaskočený členitosťou a mnohoúrovnosťou ornamentov a vzorov, ktorých pestrosť a početnosť znemožní uchopiť predlohu ako jednoliaty harmonický celok, čím vznešenosť nahrádza krásu z jednoty.

Možno to je dôvod, prečo pri mnohých ornamentoch možno nachádzať relatívne jednoduché tvary a ich početné repetície, neraz dokonca s logikou fraktálov opakujúcich vzory na rozličných úrovniach, tak, aby percipient dokázať uchopiť komplexnosť predlohy, ale zároveň aj so schopnosťou retencie zmysluplného harmonického celku. Escherove či Penroseho donekonečna sa opakujúce obrazce percepčne organizujeme do takých celkov, ktoré sú zmysluplné a ktoré dokážeme uchopovať. Pokiaľ sa nám to nedarí, krásu obsiahnutú v harmónii celku a časti nám uniká.

Pekným príkladom vzťahu jednoduchosti a komplexnosti môže byť zdobenie tváre. V mnohých kultúrach je ženská tvár zahalená a jediné čo možno uzrieť, sú oči. Ich jedinečnosť je tým zvýraznená a ich krásu nie je rušená ostatnými časťami tváre,

ktoré by ich mohli utláčať do úzadia. Často sú preto ešte zvýrazňované maľovaním či iným zdobením. Naopak, množstvo kultúr a ich častí využíva pri rôznych príležitostiach maľovanie či skrášľovanie tváre tetovaním. Ak je však tetovanie príliš zložité a rozsiahle, krásu jednotlivých prvkov tváre zaniká, či dokonca neumožňuje percipientovi ju vôbec vnímať. Preto sa často možno stretnúť len s čiastočným tetovaním – s použitím ornamentu či motívov, ktorý zaujme, či zvýrazní celok, nie nás celkom pohltí. Šperky a či iné doplnky sú toho jasným dôkazom.

Otázka vzťahu jednoduchosti a komplexnosti je značne relatívna, no ukazuje sa, že väčšina z nás preferuje stredný stupeň komplexnosti jednotlivých podnetov. Veľmi jednoduché podnety sú nenápadné, či dokonca nudné. Symetria, ktorá je najjednoduchším algoritmom replikácie je preto sice často považovaná za nevyhnutnú podmienku pre vnímanie krásy, ale nie za dostatočnú, práve preto, že jej odhalenie je prirýchle a pozornosť percipienta tak podnet ďalej nepúta. Pre zatraktívnenie jednoduchej symetrie preto neraz možno prikročiť k jej jednoduchému narušeniu (napríklad znamienkom krásy, ako tomu bolo v baroku – fluktuáčná asymetria; či cielenému narušaniu osovej symetrie celku v rokoku), čo nútí percipienta posudzovať predlohu aj v iných rámcoch a úrovniach. Naopak, pri veľmi zložitých objektoch a kontextoch môže strohost, jednoduchosť pôsobiť osviežujúco a nádherne v kontraste s intelektuálnou a zmyslovou presýtenosťou okolia. Jednoduchosť a zložitosť tak predstavujú kontinuum, ktorého póly nás odkazujú k inej dimenzii estetickej skúsenosti, a sice k pozornosti a aktivite vedomia.

2.6 Aktivita a pasivita

Dimenzie simplicity, exkluzivity aj príjemnosti odkazujú k vôli a aktivite. Zdá sa, že pre značnú časť krásnych objektov platí, že upútavajú našu pozornosť. Krásne veci sú príťažlivé, a to do takej miery, že nie lenže na ne zameriavame svoju pozornosť, ale s nimi trávime (asoň v mysli) aj nemalý čas. Vyhľadávame ich prítomnosť, radi ich percipujeme. Z tohto dôvodu sa nimi často obkolesujeme, zadovažujeme si ich a túžime ich vlastniť, ale aspoň inak s nimi pobývať. A to nielen fyzicky, ale rovnako, ak nie viac, aspoň v predstavách.

Semir Zeki upozornil na to, že neuronálna aktivita spojená s vnímaním krásnych objektom aktivizuje centrá odplaty a osobitne oblasť v medio orbitofrontálnom kortexe, kým vnímanie škaredých objektov možno spájať s aktivitou v senzomotorickom kortexe (Ishizu & Zeki, 2013, 2014). Tým možno vysvetliť, že pri vnímaní krásnych objektov máme tendencie kontemplovať a upriamovalať na ne sústredenú pozornosť, kým pri škaredých podnetoch reagujeme odvrátením ich pôsobenia na naše zmysly, hoci vo vedomí môžu stále pretrvávať. Krásne aj škaredé veci upriamujú našu pozornosť, ale akoby opačným smerom

a k odlišným aktivitám. Kým pre krásne často platí, že ich aktívne vyhľadávame a chceme sa k nim čo najviac a nerušene priblížiť a čo najväčšmi pri nich zotrvať, pri škaredých podnetoch sa snažíme ich bezprostredne pôsobenie obmedziť. Paradoxne však vedomie ošklivosti objekt z myseľ celkom nevytesňuje (hoci by sme to často radi), ale naopak, práve pre nepríjemnosť jeho pôsobenia, musíme neustále na neho myslieť (nie len v zmysle traumatizovania), ale najmä preto, aby sme sa jeho ďalšiemu potenciálnemu pôsobeniu mohli vyhnúť. Intenzívne krásne a intenzívne škaredé veci sa tak podobným spôsobom zmocňujú našej myseľ.

Pôsobenie krásy často je alebo aspoň môže byť pre jej príjemnosť späť s jej aktívnym vyhľadávaním. Po krásnych objektoch túžime, vyhľadávame ich a niekedy ich chceme vlastniť a posesívne žiarlime. Krásu tak nabáda k aktivite, náruživosti a túžbe, je skutočne atraktívna.

Podstatou atraktivity je pritom jej schopnosť mentálnej gravitácie. Podobne ako v prípade gravitačnej príťažlivosti, aj tu krásu príťahuje zmysly a myseľ, a to tým viac, čím viac sa nachádzame v akčnom rádiuse jej možného pôsobenia. Podobne je to aj s odpudivou silou škaredosti, ktorá je priamo úmerná intenzite jej pôsobenia a subjektívnej pravdepodobnosti jej výskytu.

Nie každá krása však nutne vedie k neutíchačnej aktivite, túžbe a horúčkovitej snahe po jej čo najintenzívnejšom vnímaní. Existuje aj opačný efekt pôsobenia krásnych objektov. Ten je späť skôr s tichým a pokojným kontemplovaním podnetov, s harmoniou a spokojnosťou. Niektoré estetické skúsenosti sú harmonizujúce, nabádajú spomalosť, stíšiť sa nehýbať, vedú skôr k pocitu bezčasia a večnosti a k odzbrojujúcim, respektívne dávajúcim zabudnúť na snahu po priblížení sa a vlastnení. Ba dokonca nechávajú zaniknúť aj celý subjekt, ktorý nazerá. Práve tak opisuje Schopenhauer pôsobenie hudby, práve tak možno vnímať viaceru estetických skúseností v galériach či v prírode, kedy percipient pocitíuje úplné pohltenie scenériou či predstavou a zabudnutie, či rozplynutie subjektu. Tento druh skúsenosti pripomína mystickú či hlbokú náboženskú skúsenosť, hovorí o splynutí s objektom, a možno práve preto má estetická skúsenosť skutočne blízko k istej mystike. Je len veľmi ľahko vyjadriteľná, pretvárajúca, obohacujúca a je dotykom s niečím veľmi hodnotným, prekračujúcim nás.

Pôsobenie krásnych podnetov na našu myseľ však nemusí viest (a často ani nevedie) k rovnakým reakciám, a to dokonca nielen u rozličných subjektov, ale aj u toho istého v odlišnom čase. Túžba a dychtivosť po kráske sa často stupňujú a možno ich na chvíľu utísť, ale aj zvrhnúť do averzie, pokial na nás pôsobí podnet pridľho alebo prislino. Naopak, pokoj a harmonia môžu byť niečím, čo daný podnet obvykle vyvoláva, ale s postupom času a zmenou okolností túto schopnosť objekty môžu strácať, alebo práve pre ich príjemnosť, môžu subjekt nabádať ich

vyhľadávať. A tak paradoxne, práve snaha o onú bezčasnosť a harmóniu môžu viesť k dychtivej závislosti od intenzívnej estetickej skúsenosti, podobne ako pri iných skúsenostiach.

3 Záver

Predkladaný model nepochybne nepostihuje všetky dimenzie krásy. To ani nebolo jeho ambíciou. To, čo by mohol ponúkať, je zaujímavý nástroj na mapovanie rôznych estetických zážitkov a ich obsahov prostredníctvom vytvorenia sémantického priestoru estetickej skúsenosti. Jednotlivé estetické skúsenosti sa nepochybne odlišujú, no práve prostredníctvom ich priestorových koordinát v ich sémantickom priestore im možno lepšie porozumieť a možno aj interpersonálne ich porovnávať. Na to by bolo samozrejme potrebné presnejšie poznáť vzťahy medzi jednotlivými dimenziami, ale aj to, či predstavujú lineárne rastúce kontinuum, alebo je ich postupnosť na ich osiach napr. geometrická.

Uvedený model, aj vo svojich hrubých črtách, ponúka porozumenie a vysvetlenie podobností a príbuznosti rôznych aspektov celkom odlišných skúseností. Vzhľadom na to, že krása a škaredosť sa nachádzajú sice na odlišných pôloch dimenzie pôžitku, obe však pôsobia podobne atraktívne, pretože sa môžu nachádzať na rovnakých miestach opísaných dimenziou aktivity. Obe pripáhajú našu pozornosť a možno práve to vysvetľuje, prečo sa nám páčia rozprávky, napr. bratov Grimmovcov, kde sa vyskytujú desivé scény, príšery a navonok nesmierne odpudivé bytosť. Podobne ako je to z pohľadom Medúzy, hrôza a škaredosť môže upútať našu pozornosť rovnako ako krása, hoci s odlišným pôžitkom.

Rovnako tak možno rozlišovať medzi vzrušujúcou atraktivitou tváre či ženskej postavy a pokojom nabítou a harmonizujúcou krásou dieťaťa. Kým v prvom prípade možno pociťovať príjemné páčenie sa spolu s pudním byť v blízkosti či dokonca v príazni atraktívnej bytosť, druhý príklad nás vedie skôr k spokojnosti a utlmeniu konania. Výsledkom milého pohľadu dieťaťa je skôr stísenie a potreba ochrany, než snaha o dosiahnutie objektu nazerania. Rôzna intenzita túžby, vzrušenia a žiadostivosti charakterizuje rôzne typy estetickej skúsenosti, ktoré môžu skórovať rovnako v iných dimenziách estetickej skúsenosti, a práve to ich umožňuje odlišovať.

Jedným z klúčových aspektov estetickej skúsenosti je miera a intenzita páčenia sa. Estetický súd môže oceňovať kvalitu elaborácie, originalitu a iné vlastnosti, avšak nemusí byť sprevádzaný s intenzívnym osobným páčením sa. Krásu je však späť so záľubou a páčením sa, s istou mierou slasti a pôžitku či príjemného vnímania. Páčiť sa nám však môže aj niečo, čo je samo v inom kontexte a či intenzite nepríjemné (masochistická slast), ale v danej situácii to vnímame ako príjemné. Otázka bolesti a rozkoše

predstavuje veľmi citlivú a sofistikovanú tematiku, ktorá len dokumentuje, že vzťah medzi príjemným a nepríjemným je veľmi dynamický a komplikovaný. Preto je komplikované posudzovať pocit z vnímania krásy podľa jeho fenomenálnych charakteristik a intenzity. Napriek tomu, práve doplnanie sa jednotlivých dimenzií krásy nám umožní lepšie sa v nich orientovať a rozumieť im.

Gärdenforsov model konceptuálnych priestorov umožňuje priestorovo opísat' jednotlivé pocity späť s vnímaním krásy a umiestniť ich do sémantického priestoru, ktorý sýti obsahy pojmu krásy. Je zrejmé, že pojem krásy tvorí veľmi spletitý a multidimensionálny priestor, ktorého úplné vymedzenie a opisanie nebude iste ľahké.

Poděkovanie

Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja na základe Zmluvy č. APVV-15-0294.

Literatúra

- Démuth, A. (2016): Tri historické matematické prístupy ku skúmaniu krásy. In: *Filozofia* 71, No. 9, 759–770.
- Gärdenfors P. (2004): Conceptual Spaces as a Framework for Knowledge Representation. In: *Mind and Matter*, Vol. 2 (2), 9 – 27.
- Gärdenfors, P. (2000): *Conceptual spaces: the geometry of thought*. Cambridge, Mass.: MIT Press.
- Gärdenfors, P. (2014): *Geometry of meaning: semantics based on conceptual spaces*. Cambridge, Mass.: MIT Press.
- Gärdenfors, P., Warglien, M. (2013): The development of semantic space for pointing and verbal communication. In Hudson, J., Magnusson, U., Paradis, C. (eds). *Conceptual Spaces and the Construal of Spatial Meaning: Empirical Evidence from Human Communication*, Cambridge University Press, Cambridge, 29–42.
- Graham, G. (2014): Aesthetics as a Normative Science. Royal Institute of Philosophy. Supplement, 75, 249–264.
- Guyer, P. (1979): *Kant and the Claims of Taste*, Cambridge, Mass.: Harvard University Press.
- Hume, D. (2008): *Selected Essays*. Oxford World's Classics, New York: Oxford University Press.
- Ishizu, T., Zeki, S. (2013): The brain's specialized systems for aesthetic and perceptual judgment. In:

European Journal of Neuroscience, vol. 33, 1413–1420.

Ishizu, T., Zeki, S. (2014): A neurobiological enquiry into the origins of our experience of the sublime and beautiful. In: *Frontiers in Human Neuroscience*, vol. 8, 1–10.

Kant, I. (1975): *Kritika soudnosti*. Praha: Odeon.

Kawabata, H., Zeki, S. (2004): Neural Correlates of Beauty. In: *Journal of Neurophysiology*, vol. 91, 1669–1705.

Levinson, J. (2011): "Beauty Is Not One: The Irreducible Variety of Visual Beauty", in P. Goldie & E. Schellekens, eds., *The Aesthetic Mind*. Oxford UP.

Livio, M. (2006): *Zlatý řez. Příběh fí, nejpodivuhodnéjšího čísla na světě*. Praha: Argo/Dokořán.

McMahon, J. (2001): Beauty. In Gaut, B. and Lopes, D. (eds). *The Routledge Companion to Aesthetics* (London: Routledge): 227–38.

Schiller, F. (1992): O půvabu a důstojnosti. In. *Výbor z filozofických spisů*. Prekl. Demel, F – Ozarčuk, I. Praha, Svoboda – Libertas, 73–108.

Schopenhauer, A. (2011): *The World as Will and Representation*, volume I, Cambridge: Cambridge University Press.

Reid, T. (1969): *Essays on the Intellectual Powers of Man*. Cambridge, Mass.: MIT Press.

Redukcia kognitívnych odchýlok u finančných profesionálov prostredníctvom kontrafaktov

Katarína Dudeková, Lenka Kostovičová, Matúš Konečný

Ústav experimentálnej psychológie CSPV SAV

Dúbravská cesta 9, 84101 Bratislava

katarina.dudekova@savba.sk, lenka.kostovicova@savba.sk, matus.konecny@savba.sk

Abstrakt

V sérii výskumov uvažovanie o situácii formou „čo by bolo keby“ redukovalo niektoré omyly v posudzovaní a rozhodovaní. My sme sa vo výskume zamerali na štyri kognitívne odchýlky: sklon k seba-potvrzdzaniu, k udržaniu status-quo, k zanedbaniu menovateľa a na averziu k strate. Experimentu sa zúčastnilo 52 finančných profesionálov. Navodením kontrafaktových myšlienok realizovali racionálnejšie finančné voľby a menej preferovali pôvodnú, nevýhodnú, investičnú alternatívu. Použitá intervencia nepomohla znížiť opomenutie menovateľa pri posudzovaní štatistických údajov. Tendencie uprednostniť možnosti v prospech pôvodného presvedčenia sa finančníci dopúšťali do minimálnej miery. Kontrafakty sa zdajú byť z časti úspešnou stratégiou tzv. „debiasingu“.

1 Kognitívne odchýlky vo finančnej doméne a kontrafaktový priming

Kognitívne odchýlky predstavujú bariéry optimálnych úsudkov a rozhodnutí v bežnom aj profesionálnom živote. V dnešnom komplexnom svete sa čím ďalej tým častejšie dostávame do situácií, v ktorých podľahnutie týmto odchýlkám môže priniesť závažné následky. Práve preto je odolnosť voči kognitívnym nástrahám predpokladom úspešného fungovania jednotlivcov. Profesionáli z oblasti financií sú jednou z tých populácií, u ktorých sú vysoké nároky na racionálnosť rozhodnutí aj z dôvodu, že ich suboptimálne rozhodnutia sú častokrát nákladné. Ako teda môžeme zabezpečiť, aby odchýlkam podliehali v menšej miere?

Výskum v oblasti kognitívneho zlyhávania sa postupne posúva od identifikovania chýb a omylov našej mysele k hľadaniu možností ich redukcie. Štúdie však zatiaľ neviedli k jednoznačným záverom či dostatočnému úspechu v predchádzaní kognitívnym odchýlkam, tzv. "debiasingu" (Larrick, 2004). Zistenia predošlých výskumov (Galinsky a Moskowitz, 2000; Kray a Galinsky, 2003; Strachanovej, 2017) však naznačujú, že úspešnou stratégiou pri určitých odchýlkach môže byť navodenie tzv. „kontrafaktového myslenia“.

Každý z nás sa už niekedy zaoberal kontrafaktovými myšlienkami „čo by bolo keby“. Kontrafaktové myslenie je mentálne simulovanie možných alternatív minulosti, ktoré sú v rozpoze s reálnou skutočnosťou. Horšie, či lepšie alternatívy k minulosti (kontrafakty nadol, či nahor) formulujeme zväčša spontánne a nezámerne (Ruiselová, 2009). V prípade, že kontrafaktové uvažovanie realizujeme cielene a systematicky, stáva sa významnou súčasťou kontextovej analýzy situácie a voľby stratégie (Hendrickson, 2008).

My sme sa rozhodli overiť navodenie kontrafaktového nastavenia mysele ako debiasingovú stratégiu u profesionálov z oblasti financií. Vytvorili sme batériu úloh pokrývajúcu štyri kognitívne omyly, ktoré môžu limitovať ich úspešnosť pri výkone svojej profesie.

1.1 Kontrafaktový priming

Galinsky a Moskowitz (2000) formou špecifického nastavenia mysele pomocou kontrafaktového primingu (prezentovanie udalosti, ktorá nenastala), dosiahli u participantov spontánne generovanie kontrafaktových myšlienok. Následne, v nesúvisiacich úlohách, participanti ich štúdie neposudzovali údaje skreslene smerom k potvrzdzaniu predpokladu, no brali do úvahy i protichodné údaje, a boli úspešnejší v rámci riešenia problému na kreativitu. Experimentálna manipulácia však zhoršila výkon v logickej úlohe na deduktívne uvažovanie. Prostredníctvom podobného dizajnu Kray a Galinsky (2003) zistili, že aktivácia kontrafaktového nastavenia mysele zlepšuje výkon skupín v úlohách, kde je zvažovanie iných alternatív a protichodných informácií kľúčové. Skupiny boli vďaka kontrafaktovému mysleniu menej náhylné podľahnúť sklonu k seba-potvrzdzaniu (*confirmation bias*). Do tretice, v štúdiu Strachanovej (2017) kontrafaktový priming znižil nielen tendenciu utvrdzovať sa v pôvodnom presvedčení, ale aj sklon pripisovať príčiny správania iných ľudí osobnostným, a nie situačným, faktorom (tzv. atribučný omyl).

1.2 Sklon k seba-potvrdzovaniu

Kognitívne vysvetlenie skonu k seba-potvrdzovaniu - vyhľadávaniu, interpretovaniu a uprednostňovaniu informácií spôsobom, ktorý potvrdzuje už existujúce presvedčenie jednotlivca (Baron, 2000) - je založené na limitovanej schopnosti ľudí zvládať komplexné úlohy a na mentálnych skratkách, ktoré ľudia následne používajú (Friedrich, 1993). Dokonca aj malá zmena vo formulácii môže ovplyvniť vyhľadávanie argumentov a formulovanie záverov. V oblasti financií môže viesť investorov k tomu, aby preceňovali svoje schopnosti a ignorovali skutočnosť, že ich stratégie môžu byť stratové (Pompian, 2012).

1.3 Sklon k udržaniu status quo

Skon k udržaniu pôvodnej pozície predstavuje u ľudí preferenciu súčasného stavu a zotrvanie v možnosti, prezentovanej pred situáciou voľby. Status quo pôsobí ako referenčný bod, od ktorého je každá zmena vnímaná ako zisk alebo strata. Samuelson a Zeckhauser (1988) sledovali výskyt tohto javu v rámci finančnej oblasti a konštatujú, že ľudia majú tendenciu bojovať proti zmene, ak výhody dostatočne neprevyšujú riziko. Čo sa týka kognitívnych limitov rozhodovania, alternatívny ponúkajúce status quo sú menej náročné na kognitívne úsilie (Eidelman a spol., 2009).

1.4 Zanedbanie menovateľa

Výber odchýlky vychádzal z predpokladu, že zamestnanci vo finančnej sfére v značnej miere pracujú s číslami. Pri prezentovaní číselných formátov dochádza k situáciám, kedy ten istý numerický údaj, prezentovaný dvoma rôznymi spôsobmi (napríklad 5 zo 100 verus 45 z 900), je posudzovaný ináč. Ľudia nedostatočným venovaním pozornosti menovateľovi nezohľadňujú základný výskyt (Reyna a Brainerd, 2008). Ako tvrdia Alonso a Fernández-Berrocal (2003), dôvodom môže byť lepšie uchopenie početnosti než vzťahov medzi číslami (*numerosity effect*), prípadne nízkych čísel než veľkých množstiev (*small numbers effect*). Porozumenie pravdepodobnostiam s väčším menovateľom tak môže byť pre ľudí náročnejšie (Garcia-Retamero a Galesic, 2011). Tento fenomén a jeho výskum má už dlhšiu tradíciu a v literatúre sa zvykne označovať ako zanedbanie menovateľa (*denominator neglect*; Ballová Mikušková, 2015).

1.5 Averzia k strate

Z oblasti finančných volieb sme do výskumu zahrnuli odchýlku od normatívnych volieb v situácii rizika, averziu k strate. Do psychológie usudzovania a rozhodovania ju uviedli Kahneman a Tversky (1979). Často skloňovaný fenomén je jedným z parametrov ich prospektovej teórie. Vypovedá o tom, že strata istej hodnoty je pre ľudí závažnejšia ("viac bolí") než výhra

rovnamej výšky (ktorá "menej teší"). V rizikových voľbách sa averzia k stratám prejavuje ako neochota hrať hru napriek jej pozitívnej očakávanej hodnote (t.j. možná výhra je vyšia než potenciálna prehra).

1.6 Výskumný cieľ a hypotézy

V našej experimentálnej štúdii sme chceli overiť vplyv navodeného kontrafaktového myšlenia na vybrané kognitívne odchýlky. Zaujímalo nás, či aktivácia myšlenia "čo by bolo keby" pomôže znížiť sklon k seba-potvrdzovaniu, udržaniu status-quo, zanedbaniu menovateľa a averziu k strate. Vzhľadom k profesionálnej skúsenosti participantov sme predpokladali, že numerické kognitívne úlohy pre nich nebudú príliš náročné, čím vznikne priestor pre efekt indukcie kontrafaktového myšlenia.

2 Metódy

2.1 Participanti

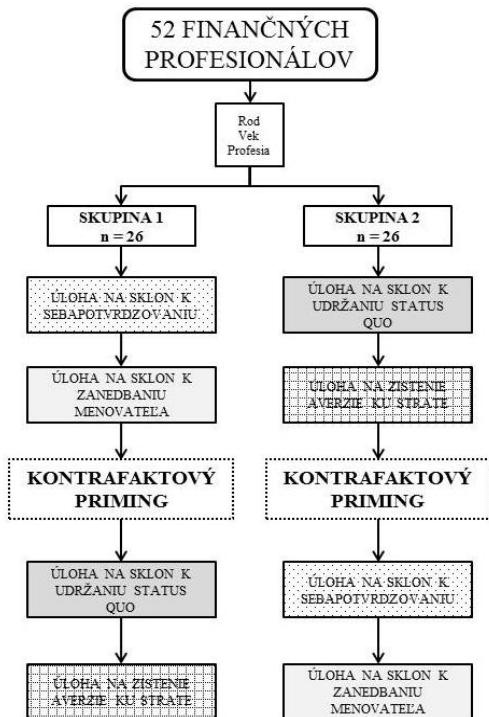
Výskumu sa zúčastnilo 54 ľudí, pričom dvaja z nich boli na základe nesplnenia podmienky práce vo finančnom sektore vyradení. Analyzovali sme dátu od 52 finančných profesionálov vo veku 20 až 63 rokov ($M = 38,7; SD = 10,8$). Výskumná vzorka bola relativne vyvážená vzhľadom na rod, kedže muži tvorili 59,6% ($n = 31$) a ženy 40,4% ($n = 21$) výskumného súboru. Medzi profesiami participantov sa objavovali najčastejšie nasledovné: ekonóm, finančný sprostredkovateľ, finančný poradca, finančný agent, hypoteckárny finančný konzultant, bankár, účtovník či hypoteckárny špecialista.

2.2 Dizajn

V rámci medzi-subjektového experimentálneho dizajnu boli participanti náhodne pridelení do jednej z dvoch skupín, ktoré sa líšili v poradí prezentovaných úloh. Dve úlohy zodpovedali finančníci pred a dve po manipulácii. Skupiny si navzájom boli kontrolou a experimentálnou skupinou, tzn. že úlohy, ktoré skupina 1 riešila pred manipuláciou, riešila skupina 2 po manipulácii a naopak. Experimentálna manipulácia spočívala v scenárii s kontrafaktami rôzneho smeru (kontrafaktový priming). Presný dizajn výskumu zobrazuje Obr. 1.

2.3 Materiály a procedúra

Participantom bol elektronickou formou distribuovaný dotazník so základnou inštrukciou. Po zodpovedaní demografických otázok boli priradení do jednej z experimentálnych skupín. Všetky úlohy, týkajúce sa sledovaných kognitívnych omylov, boli prispôsobené pracovnej (finančnej) doméne participantov.



Obr. 1: Dizajn experimentu

Úloha na sklon k seba-potvrdzovaniu (inšpirovaná štúdiou Galinsky a Moskowitz, 2000) sa týkala investičného poradenstva. Participanti si prečítali scenár: „Kamarátku Vás požiada, či by ste jej kolegovi poradili ohľadne finančnej investície. Považujete za potrebné zistiť, aký je jeho postoj k riziku, či sa mu radšej vyhýba alebo je voči nemu tolerantný. Na základe kamarátkiného opisu sa domnievate, že dotyčný je skôr typ, ktorý riziko akceptuje, možno i vyhľadáva. V rámci neformálneho rozhovoru si chcete predbežne overiť, či ste ho dobre odhadli. Pripravíte si zoznam otázok, no pre krátkosť času sa mu rozhodnete položiť len 5 z nich. Ktoré by to boli?“

Štyri ponúknuté otázky boli v zhode s predpokladom, že kamarátkin kolega je tolerantný k riziku, napr. „Vyhľadávaš adrenalínové aktivity?“ či „Riadiš sa v niektorých situáciách heslom risk je zisk?“ Ďalšie štyri, naopak, odporovali predpokladu, napr. „Povedali by o Tebe Tvoji priatelia, že si opatrný a neriskuješ?“ alebo „Vyhýbaš sa impulzívnym nákupom?“. Do súboru sme zakomponovali aj neutrálnu otázku, konkrétnie: „Si spoločenský typ?“ V prípade, že finančný profesionál volil viac otázok v zhode s pôvodným predpokladom (kamarátkin kolega toleruje riziko) a menej otázok v protiklade s predpokladom (kolega sa riziku vyhýba), dopustil sa sklonu k seba-potvrdzovaniu. Ak zvolil napríklad 3 otázky v zhode s predpokladom 2 opačné, jeho skóre je +1. Takto mohol dosiahnuť skóre -4 až +4.

Úloha na sklon k udržaniu status quo (podľa Samuelson a Zeckhauser, 1988) znala: „Predstavte si,

že ste zdedili po vzdialenej príbuznej portfólio, ktorého prevažná časť je investovaná do stredne-rizikovej Spoločnosti A. Zvažujete či ostanete pri tejto možnosti alebo sa rozhodnete investovať inam“. Participanti si vyberali z troch možností, pričom jedna bola zároveň pôvodnou alternatívou, ktorá bola z hľadiska investovania najmenej vhodná:

"a) Pôvodná možnosť – stredne-riziková Spoločnosť A. Prognóza pre nasledujúce 3 roky: šanca, že Vaše akcie stúpnu o 30% je 0,5, šanca, že ich hodnota ostane rovnaká je 0,2 a riziko, že ich hodnota klesne o 20% je 0,3.

b) Investícia do vysoko-rizikovej Spoločnosti B. Prognóza pre nasledujúce 3 roky: šanca, že sa hodnota Vašich akcií zdvojnásobí je 0,4, šanca, že ich hodnota ostane rovnaká je 0,3 a riziko, že ich hodnota klesne o 40% je 0,3.

c) Investícia do nízko-rizikovej Spoločnosti C. Prognóza pre nasledujúce 3 roky: šanca, že Vaše akcie stúpnu o 20% je 0,7, šanca, že ich hodnota ostane rovnaká je 0,2 a riziko, že ich hodnota klesne o 5% je 0,1".

Preferovanie pôvodnej (aj keď pre investície značne nepriaznivej) možnosti hovorí o skлонke k udržaniu status quo.

Posudzovanie štatistických údajov a **zanedbanie ich menovateľa** sme zisťovali pomocou nasledujúcej úlohy: „Nová mobilná aplikácia istej banky má pomáhať klientom spravovať svoje financie. Banka sa rozhodla zrealizovať výskum na klientoch, ktorí v prvom kole dotazníkov vyjadrili nespokojnosť so spravovaním svojich financií. Po roku, v priebehu ktorého mali klienti možnosť používať aplikáciu, vyplnili ďalší dotazník. Jeho vyhodnotenie ukázalo nasledovné. 90 zo 600 ľudí, ktorí aplikáciu v priebehu roka nepoužívali, vyjadrilo opakovane nespokojnosť so spravovaním svojich financií. 45 z 200 ľudí, ktorí aplikáciu dlhodobo používali, vyjadrilo opakovane nespokojnosť so spravovaním svojich financií. Aká prospešná bola mobilná aplikácia?“.

Participanti sa vyjadrovali na sedembodovej Likertovej škále s koncovými bodmi 1 – *Nebola prospešná* až 7 – *Bola veľmi prospešná*. Racionálnej voľbou bola odpoveď 1, vyššie číslo vyjadruje väčší sklon k zanedbaniu menovateľa.

Posledným použitým problémom bola úloha zistujúca participantovu **averziu k strate**. Inštrukcia: „Predstavte si možnosť zahrať si 6 hier, v ktorých sa bude hádzať mincou. Ak padne v danej hre rub, vyhráte 6 €; ak v nej padne lice, prehráte sumu od 2 do 7 €.“ Úlohou je rozhodnúť sa, či človek konkrétnu hru odmietne alebo prijme. Averzia k strate predstavuje odmietnutie zúčastniť sa hry v prípade, že možnosť výhry je vyššia ako možnosť prehry (Bačová a Jusková, 2009). Veľkosť averzie k strate vyjadruje počet odmietnutých hier (0 až 7).

Kontrafaktový priming

Scenár sme zostavili podľa vzoru východiskovej štúdie (Galinsky a Moskowitz, 2000), pričom sme ho doplnili o viaceré kontrafaktové podnety. Konkrétnie znenie: "Janka je na koncerte svojej oblúbenej kapely. V koncertnej hale je voľné sedenie. Jeden z organizátorov nečakane oznamí, že šťastlivec, ktorý bude sedieť na vyžrebovanom čísle sedadla, vyhrá dovolenkú na Havaji. Janka chce mať lepší výhľad a presadne si do jedného z predných radov. Zistí, že na vedľajšom mieste sedí jej kamarátka z detstva, ktorú dlho nevidela, veľmi sa poteší a dajú sa do reči. Dievčatá si koncert vychutnajú, no zážitok im trochu pokazí pán pred nimi, ktorý koncert neustále fotí. Na záver ohľásia číslo výherného sedadla. Výhercom sa stáva pán sediaci na mieste, z ktorého Janka odišla. Janka sa cestou z koncertu zamýšla, čo by získala keby si nepresadla, a o čo by naopak prišla, keby si presadla." V závere sme overovali, či si príbeh (v súlade s inštrukciou) participanti prečítali pozorne a do analýzy postúpili iba tí, ktorí zvolili možnosť "áno".

3 Výsledky

3.1 Sklon k seba-potvrdzovaniu

Participanti si zo štyroch možností na potvrdenie predpokladu (tolerancia k riziku) vyberali 1 až 3 otázky tak v kontrolnej ("KS"; $Mdn = 2,00$; $IQR = 1,00$), ako i experimentálnej skupine ("ES"; $Mdn = 2,00$; $IQR = 1,00$). Členovia KS volili o niečo viac otázok v prospech predpokladu ($MRank = 28,81$) ako členovia ES ($MRank = 24,19$), no išlo o malý a nevýznamný rozdiel, $M-W U = 278,00$; $p = 0,213$; $rm = 0,17$.

Participanti v KS si zároveň vyžiadali o niečo menej otázok v protiklade s predpokladom (averzia k riziku) ($Min = 1$; $Max = 4$; $Mdn = 2,00$; $IQR = 1,00$; $MRank = 25,13$) v porovnaní s ES ($Min = 0$; $Max = 3$; $Mdn = 3,00$; $IQR = 1,00$; $MRank = 27,87$), no rozdiel, opäťovne, neboli signifikantný, $M-W U = 302,50$; $p = 0,476$; $rm = 0,10$.

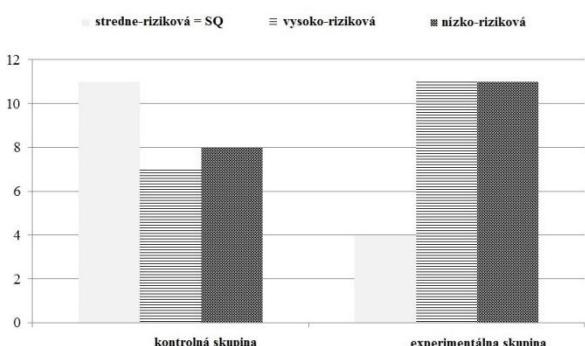
Taktiež porovnanie sumárneho skóre úrovne sklonu k seba-potvrdzovaniu (rozdiel v počte otázok v prospech a v neprospech predpokladu, s možným rozpätím -4 až +4) neprinieslo signifikantné výsledky. U finančníkov v KS sa hodnoty pohybovali medzi -3 až 2 ($Mdn = 0,00$; $IQR = 2,00$; $MRank = 28,85$) a v ES medzi -2 až 3 ($Mdn = -1,00$; $IQR = 2,00$; $MRank = 24,15$), $M-W U = 277,00$; $p = 0,233$; $rm = 0,17$.

Overili sme tiež rodové rozdiely a vzťah tendencie k seba-potvrdzovaniu s vekom. Muži ($n = 31$; $Mdn = 3,00$; $IQR = 1,00$; $MRank = 30,11$) volili významne viac otázok v neprospech predpokladu ako ženy ($n = 21$; $Mdn = 2,00$; $IQR = 1,00$; $MRank = 21,17$), $M-W U = 213,50$; $p = 0,022$; $rm = 0,15$. Vek pozitívne koreloval s počtom otázok odporúčiacich predpokladu, $\rho = 0,41$; $p = 0,003$, a negatívne s počtom otázok podporujúcich predpoklad, $\rho = -0,36$; $p = 0,009$,

i s celkovým sklonom k seba-potvrdzovaniu, $\rho = -0,42$; $p = 0,002$.

3.2 Sklon k udržaniu status quo

Ako je možné vidieť na Obr. 2, v KS by pôvodnú možnosť investovať do stredne-rizikovej spoločnosti A volilo, a tým pádom udržiavalo počiatočný status quo, 43,2 % ($n = 11$) finančníkov a v rámci ES iba 15,4 % ($n = 4$) participantov, čo predstavuje významný rozdiel, $\chi^2(1) = 4,59$; $p = 0,032$; $\phi = 0,61$. Rozdiely podľa rodu a veku sme nenašli.



Obr. 2: Sklon k udržaniu status quo podľa skupiny

3.3 Sklon k zanedbávaniu menovateľa

Spomedzi 52 finančných profesionálov nezanedbalia informáciu v menovateľovi približne jedna tretina (30,8%; $n = 16$). Z počtu 16 sa v KS nachádzalo 5 participantov, v ES ich bolo 11 (čo naznačuje zlepšenie vplyvom kontrafaktových myšlienok). Finanční profesionáli však, celkovo, podliehali tejto odchýlke iba v malej miere ($Mdn = 1,00$; $IQR = 2,00$). V rámci KS sa ľudia častejšie ($Mdn = 2,00$; $IQR = 1,00$) priklonili k tomu, že aplikácia je prospešná, čím vo viacerých prípadoch opomenuli informáciu v štatistickom údaji (ES: $Mdn = 1,00$; $IQR = 2,00$). Rozdiel nedosiahol úroveň štatistickej signifikancie, no hladina vecnej významnosti dosiahla mierne hodnoty, $M-W U = 256,00$; $p = 0,122$; $rm = 0,22$.

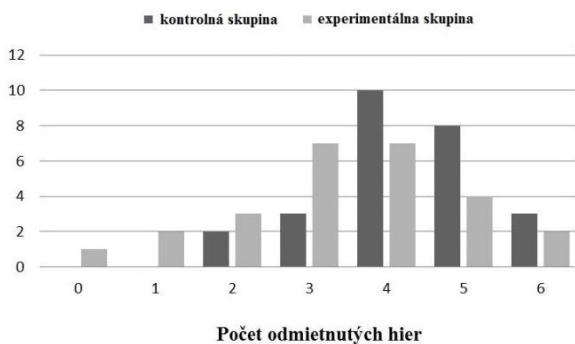
Vek so zanedbaním menovateľa nesúvisel. No muži - finančníci, narodzeli od žien, vnímali štatistické údaje v úlohe na sklon k zanedbaniu menovateľa menej skreslene. Z celkového počtu správnych odpovedí bola väčšina (75%) mužských. Rozdiel medzi mužmi ($Mdn = 1,00$; $IQR = 2,00$; $MRank = 22,89$) a ženami bol signifikantný ($Mdn = 2,00$; $IQR = 2,50$; $MRank = 31,83$), $M-W U = 213,50$; $p = 0,031$; $rm = 0,30$).

3.4 Averzia k strate

Na Obr. 3 je možné vidieť distribúcie počtu odmietnutých hier. V rámci kontrolnej skupiny participanti odmietali hrať v priemere vyšší počet hier ($M = 4,27$; $SD = 1,08$) ako v experimentálnej skupine

($M = 3,42$; $SD = 1,50$), pričom išlo o signifikantný rozdiel, $t(50) = -2,33$; $p = 0,024$, $d = 0,66$.

Vzťah veku a averzie k strate bol nevýznamný. Muži ($M = 3,52$; $SD = 1,36$) však odmietli výrazne menej hier ako ženy ($M = 4,33$; $SD = 1,24$), $t(50) = -2,20$; $p = 0,032$, $d = 0,55$.



Obr. 3: Distribúcia odmietaných hier podľa skupiny

4 Diskusia

Hoci sa človek môže cítiť dostatočne zručný v riešení problémov - vedľ predsa zvažuje a rozhoduje sa mnohokrát denne - často sú produktom systematické a predvídateľné odchýlky od plnej racionálnosti, čo potvrdilo i mnoho dekád výskumu (pre prehľad napr. Kahneman, 2011). Skúmanie kognitívnych skreslení sa zdá byť presýtené množstvom štúdií, poukazujúcich na zlyhania v ľudskom uvažovaní. Vyhľadávaná otázka, nakoľko je potrebné a prospešné zaoberať sa všetkými možnosťami ako isté procesy nefungujú, respektíve sú „pokazené“ (výskum *bases*) alebo zacieliť svoju pozornosť na možnosti „opravy“ (*debiasing*)? Preto bolo cieľom nášho výskumu overiť účinnosť metódy redukcie vybraných kognitívnych odchýlok, ktorá sa môžu vyskytnúť pri práci finančných profesionálov, a zároveň v rámci rozhodovania každého z nás.

Ked' sa ocitneme v problematickej situácii, spravidla hľadáme rôzne možné riešenia. Nemalá časť našich myšlienok obvykle smeruje k zvažovaniu iného chodu udalostí, veľakrát v protiklade k skutočnosti: čo by bolo keby? Podľa predošlého výskumu (Galinsky a Moskowitz, 2000; Kray a Galinsky, 2003; Strachanová, 2017) sa práve jednou zo stratégii „debiasingu“ javí byť kontrafaktový priming, čiže zvažovanie alternatívneho vývoja udalostí. Preto i nás zaujalo, či aktivácia myslenia „čo by bolo keby“ prostredníctvom scenára, ktorého pozorné prečítanie by malo spustiť simuláciu alternatívnych udalostí, pomôže redukovať vybrane odchýlky – sklon k seba-potvrdzovaniu, udržaniu status-quo, zanedbaniu menovateľa a averziu k strate, u skupiny finančných profesionálov.

Vďaka myšlienkom „čo by bolo keby“ sa profesionáli v oblasti financií rozhodovali menej skreslene, respektíve sa dopustili kognitívnych odchýlok v nižšej miere, v dvoch prípadoch. Pri

investovaní menej preferovali pôvodnú, no nevýhodnú, možnosť, a menej sa vyhýbali možným stratám. Suma sumárum, ich finančné voľby boli viac racionálne než v prípade ľudí, ktorí pred ich učinením neboli vystavení kontrafaktovému primingu. Inak povedané, viac sa priblížili typickému predstaviteľovi normatívnych modelov rozhodovania „homo economicus“.

V protiklade s predošlými zisteniami (Galinsky a Moskowitz, 2000; Kray a Galinsky, 2003; Strachanová, 2017) však kontrafaktové nastavenie myšle neznížilo sklon utvrdzovať sa v tom, čo si človek pôvodne mysel. Navodenie kontrafaktových myšlienok tiež neovplyvnilo porozumenie štatistickým údajom, a sklon k zanedbaniu ich menovateľa. Ked'že sa v prípade oboch problémov participanti dopúšťali nesprávnych odpovedí v pomerne nízkej miere, je možné, že samotná profesionálna skúsenosť poskytuje určitú úroveň odolnosti voči istým typom kognitívnych odchýlok. Interpretácia je v zhode so zisteniami predošlých výskumov (Dudeková a Kostovičová, 2016; Kostovičová a Dudeková, 2016).

Čo sa týka špecifík našej vzorky, napriek snahe oslovíť príbuzné profesie je možné, že typ pracovnej činnosti a miera finančnej skúsenosti, či práce s číslami sa mohla od profesie k profesii meniť. Taktiež sa javí, že finanční profesionáli nie sú homogénou skupinou. Vek, a najmä rod účastníkov, zohral úlohu vo väčšine kognitívnych problémov. Pribúdajúcim vekom klesala tendencia k seba-potvrdzovaniu. Muži menej zanedbávali menovateľa, boli viac ochotní pripustiť možné straty a častejšie volili možnosti v neprospech pôvodného predpokladu.

Náš výskum prináša originálne a najmä optimistické zistenia ohľadne možností redukcie kognitívnych odchýlok - averzie k strate a sklonu k udržaniu status quo - ktoré majú v praxi finančných profesionálov, no i v živote ľudí všeobecne, závažné dôsledky (napr. Kahneman, Knetsch a Thaler, 1991; Samuelson a Zeckhauser, 1988). Výzvou pre ďalší výskum, samozrejme, ostáva overenie účinnosti metódy v reálnych podmienkach a hľadanie alternatívnych stratégii tzv. "debiasingu" v situáciách, kedy kontrafaktové myslenie nie je účinné.

Vzhľadom k tomu, že kontrafaktové myslenie je prirodzenou súčasťou ľudskej myšle (Roeser, 1994), jeho prijatie za individuálnu stratégiu a cielené, systematické používanie v relevantných situáciách by mohlo pozitívne prispieť k adekvátnym úsudkom, úspešnému riešeniu problémov a optimálnym voľbám. Intervencie na báze našich zistení a ďalších výskumných iniciatív, smerujúcich k odstraňovaniu kognitívnych odchýlok, preto môžu zvýšiť kvalitu tak pracovného, ako i osobného života.

Poděkování

Príspevok vznikol za podporu Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-0361-

Literatúra

- Alonso, D. a Fernández-Berrocal, P. (2003). Irrational decisions: Attending to numbers rather than ratios. *Personality and Individual Differences*, 35(7): 1537–1547.
- Bačová, V. a Jusková, T. (2009). Risky choice and affective forecasting: Loss aversion in two culturally different student samples. *Studia Psychologica*, 51(4): 329–342.
- Ballová Mikušková, E. (2015). The denominator neglect in decision-making. *Studia psychologica*, 57(4): 255–269.
- Baron, J. (2000). *Thinking and deciding*. New York: Cambridge University Press.
- Dudeková, K. a Kostovičová, L. (2016). Čo z čoho a kto z koho? Efekt číselného vyjadrenia a odbornosť pri posudzovaní rizika v oblastiach zdravia a financií. V zborníku *Kognice a umělý život XVI*, str. 23–28.
- Eidelman, S., Crandall, C. S. a Patterson, J. (2009). The existence bias. *Journal of Personality and Social Psychology*, 97(5): 765–775.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, 100(2): 298–319.
- Galinsky, A. D. a Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristics and consideration of alternatives. *Journal of Experimental Social Psychology*, 36(4): 384–409.
- Garcia-Retamero, R. a Galesic, M. (2011). Using plausible group sizes to communicate information about medical risks. *Patient Education and Counseling*, 84(2): 245–250.
- Hendrickson, N. (2008). *Counterfactual reasoning*. A basic guide for analysts, strategists, and decision makers. The Proteus Monographs Series, 2. The National Intelligence University, U. S. Army War College.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., Knetsch, J. L. a Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, Kahneman, D. a Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–292.
- Kostovičová, L. a Dudeková, K. (2016). Finančníci – pragmatici? Konverzačný kontext a doménová špecifickosť v posudzovaní rizika a rozhodovaní. V zborníku *Kognice a umělý život XVI*, str. 85–89.
- Kray, L. J. a Galinsky, A. D. (2003). The debiasing effect of counterfactual mind-sets: Increasing the search for disconfirmatory information in group decisions. *Organizational Behavior and Human Decision Processes*, 91(1): 69–81.
- Larrick, R. P. (2004). Debiasing. V knihe *Blackwell handbook of judgment and decision making*, str. 316–337.
- Pomian, M. M. (2012). *Behavioral finance and wealth management: How to build optimal portfolios that account for investor biases*. New Jersey: Wiley.
- Reyna, V. F. a Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18(1): 89–107.
- Roes, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66(5): 805–818.
- Ruiselová, Z. (2009). Kontrafaktové myslenie. V knihe *Kontrafaktové myslenie a osobnosť*, str. 7–28.
- Samuelson, W. a Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1): 7–59.
- Strachanová, D. (2017). *Čo by bolo keby? Redukcia kognitívnych odchýlok prostredníctvom kontrafaktov*. Bratislava: Univerzita Komenského.

Ja na rozhraní filozofie a vedy

Silvia Gáliková

Katedra filozofie, Filozofická fakulta, Trnavská univerzita,
Hornopotočná 23, 918 043, Trnava
email: silvia@libris.sk

Abstrakt

Intenzívny teoretický a experimentálny výskum povahy vedomej skúsenosti človeka v posledných dekádach predznamenal znovuzrodenie záujmu o skúmanie statusu Ja. Filozofi a vedci akcentujú viacvýznamosť a komplexnosť pojmu a fenoménu Ja (Self). Autor poukazuje na konceptuálnu a metodologickú spletitosť pri skúmaní toho, čo v bežnej skúsenosti vnímame ako „vlastníka myšlienok a pocitov“, „nositeľa identity seba“, „vnútorný hlas“, „subjektívne prežívanie existencie“. Predmetom analýzy sú typy Ja (ekologické, naratívne, privátne, kognitívne), štruktúra ja (seba-poznanie, seba-kontrola) a vývin ja (proto-ja, základné ja, rozšírené ja). Základný argument tvorí téza o existencii viacúrovňovej asymetrie medzi prežívaním „vnútornej skúsenosti ja“ v každodennej skúsenosti a jej teoretickým uchopením a vysvetľovaním. Argumentácia sa opiera o výskum poškodení ja, simulovanie vedomej skúsenosti ja, prípadové štúdie. Ciel príspevku spočíva v zhodnotení legitimity konceptu Ja a jeho funkcií na základe reprezentatívnych filozofických a neurovedeckých modelov (Zahavi, Metzinger, Damasio).

1 Úvod

Prežívanie a existencia stavov Ja v kontexte každodennej skúsenosti človeka predstavuje v súčasnej filozofii myšle a vede závažný problém. Aká je štruktúra, základný konceptuálny rámec subjektívnej skúsenosti? Možno pripisať Ja vlastnú empirickú realitu alebo ide o teoretickú fikciu? Aký je vzťah medzi vedomím a Ja? Aké sú minimálne kritériá pre pripísanie Ja organizmu? Akú funkciu plní Ja v našom vedomom živote? Problémy súvisiace s odpovedami na tieto pretrvávajúce otázky sú tesne späté s terminologickým zmätkom prítomným v štúdiách myšle a vedomia. K širokej palete charakteristik Ja možno zaradiť koncepcie vychádzajúce z tradičnej definície v zmysle „mentálnej veci“, „transcendentálneho Ja“, „Ega“ až po súčasné charakteristiky v zmysle „fikcie“ alebo „neurónových procesov“. Vedci a filozofi si čoraz väčšmi uvedomujú potrebu vzájomnej kooperácie medzi disciplínami. PomySELNÉ hranice medzi tra-

dične vyprofilovanými autonómnymi vednými a klinickými disciplínami sa postupne rozplývajú a vytvárajú sa nové, explanačne silnejšie prístupy v rámci neuropsychiatrie, kognitívnej psychológie, psycholinguistiky, neuroetiky atď. Posilňovanie a rozvíjanie vzájomného medziodborového skúmania nie je ani v súčasnosti všeobecne prijímané, resp. stretáva sa s mnohými prekážkami a zdaleka nepredstavuje homogénny celok. Aj to je jeden z dôvodov, prečo sa v texte budem opierať o myšlienky a prístupy autorov s podobnou pracovnou charakteristikou predmetu skúmania a možností objektívneho, systematického štúdia a vysvetlenia.

Teoretickú platformu na ktorej sa zhodnú viacerí teoretici a vedci skúmajúci povahu vedomej skúsenosti možno v stručnosti vyjadriť nasledovnými tézami:

- mysel' a Ja sú ukotvené¹, postulovanie transcendentálneho rozumu, Ja alebo vedomia existujúcich nezávisle od telesnej štruktúry (vrátane mozgu) je kategorická chyba, mýlus alebo oboje
- skúsenosť a kognícia sú ukotvené nielen v tele, ale aj v okolitom prostredí (embedded)
- realita Ja má svoj pôvod v prežívaní vzťahu k sebe samým, k druhým a v konaní vo svete
- experimentálne, empirické a klinické poznatky tvoria neoddeliteľnú súčasť skúmania povahy Ja a vedomia

Na základe uvedeného možno konštatovať, že prívrženci rozličných teoretických prístupov chápú Ja spolu s vedomou skúsenosťou ako prirodzené javy podliehajúce vedeckej metodológií a výskumu. Ako však poukážem v nasledujúcom teste, skutočnosť je oveľa problematickejšia. V dôsledku neraz protichodných teoretických východísk skúmania nesúčich so sebou rozličné ciele skúmania, nie je vždy jasné, v čom spočíva problém skúmaného feno-

¹ V teste používam pojem *ukotvenosť* v širokom význame na zdôraznenie telesnej povahy vnútornej skúsenosti, ktorá zahŕňa viacúrovňové vnútorné procesy tela, neurónovú aktivitu mozgu ako aj kontext vzájomných vzťahov medzi telom a okolitým svetom. Z tohto dôvodu pojmy *embodied* (vtelesnený, Lakoff, 1999), *embedded* (vsadený do vonkajšieho sveta, Varela a spol., 1991) a *enactive* (zahŕňajúci neurónovú aktivitu vrátane konania jedinca, Varela, 1996) chápem ako vzájomne sa podmieňujúce.

ménu a čo sa očakáva od jeho vyriešenia a vysvetlenia. Vzhľadom na veľké množstvo predostretých modelov a teórií o povahе vedomej skúsenosti, sa zameriam na vybrané prístupy z oblasti neurovedného a kognitívneho skúmania konceptu a fenoménu ja, ktoré významne zasahujú do teoretickej diskusie. Ide o koncepcie D. Zahaviho, D. Dennetta, T. Metzingera a najmä A. Damasia.

2 Pojem a ontológia Ja

V odbornej literatúre z oblasti psychológie, psychiatrie, filozofie, kognitívnej vedy sa stretávame s množstvom rozmanitých „definícii“ Ja. Spomedzi vyše dvadsiatky používaných pojmov Ja (Self) sa filozof G. Strawson hlási ku klasickému pojmu: „ked’ hovorím o Ja ... mám na mysli „vnútorné duševné veci, vnútorné duševné prítomnosti“ (Strawson, 2010, 274). Nech je Ja čimkoľvek, tvrdí Strawson, v každom prípade ide o subjekt skúsenosti a určite nejde o ľudskú bytosť chápanú ako celok. Pôvod problému Ja nachádza v každodennej skúsenosti človeka. K univerzálnym črtám takto poňatého Ja radí tieto: a) subjekt skúsenosti, niečo, čo cíti a myslí, b) vec, c) duševná vec, d) v každom časovom úseku ide o jednu vec, e) v skúsenosti pretrvávajúca vec, f) agent, g) niečo, čo má osobnosť. Na pozadí uvedených vymedzení Ja prebiehajú aj súčasné teoretické polemiky. V diskusiách sa znova utvorila otázka podstaty Ja spätá s ego-logickými a non-ego-logickými prístupmi. Pre stúpencov ego-logických teórií je vo všeobecnosti konceptuálnou a empirickou pravdou, že akákoľvek udalosť prežívania skúsenosti v sebe nevyhnutne zahŕňa subjekt skúsenosti. Non-ego-logická teória, naopak, popiera nutnosť existencie nositeľa skúsenosti, t. j. toho, že by sa každá udalosť odohrávala pre subjekt. K prívržencom ego-logického prístupu sa radí súčasný dánsky fenomenológ D. Zahavi, podľa ktorého zodpovedanie otázky o podstate Ja vyžaduje preskúmanie štruktúry skúsenosti. Ja má svoju empirickú realitu, čo znamená, že je prepojené s perspektívou prvej osoby alebo s podaním empirického javu prvou osobou (Zahavi, 2005, 106). Ja nie je neopísateľnou transcendentálnou podmienkou ani spoločenskou či narratívnou ideou, Ja je integrálnou súčasťou nášho vedomého života s bezprostrednou empirickou realitou. Prvoradým cielom fenomenologickej analýzy je opísať a vysvetliť základný štruktúrny princíp, ktorý, ako priesitor, posilňuje podstatu toho, čo je nám dané ako obsah danej skúsenosti. S kritikou fenomenologickej konceptu Ja prichádza nemecký neurovedec T. Metzinger, podľa ktorého nemá táto predstava Ja vysvetľujúcu hodnotu a preto „za“ samotným konceptom nič nestojí. Vo svojej knihe *The Ego Tunnel* T. Metzinger sformuloval hypotézy o „mýte Ja“: „Ja neexistuje. Napriek tomu, čo si väčšina ľudí myslí, nikto nikdy neboli alebo nemal Ja“ (Metzinger, 2010, 1). Odmietnutie fenomenologic-

kého konceptu Ja ako „niečoho daného“ postulovaného a priori uvažovaním priviedlo niektorých filozofov k svojráznej myšlienke, podľa ktorej Metzinger v skutočnosti popiera existenciu Ja a vedomia. Úvahy o legitimité pojmu a následného fenoménu Ja sa nesú v znamení posudzovania vzťahu existencia - neexistencia. D. Dennett, napríklad, kritizuje charakteristiku Ja v zmysle „karteziánskeho divadla“, myšlienky o mieste, „kde všetko (subjektívne skúsenosti) prebieha“. Podľa Dennetta, „tí, ktorí sa považujú za materialistov a stále sa pritom pridŕžajú karteziánskeho divadla so všetkou jeho očarujúcou obraznosťou, sú uväznení v Descartovom materializme“, názore, podľa ktorého sa niekde v mozgu nachádza koherentná reprezentácia všetkého, čo v danej chvíli vedomie prežívame (Dennett, 1991). Vo svojej koncepcii „viacúrovňových polí“ Dennett chápe Ja ako entitu rozloženú v mozgu v čase a priestore, zahrňajúcu vedomé a nevedomé procesy mozgu zodpovedné za správanie. Aj podľa S. Blackmoreovej (2002) vedomie predstavuje zaujímavú ilúziu. Na rozdiel od svojich kritikov však netvrdí, že vedomie neexistuje: „Mám na mysli, že vedomie nie je tým, čím sa nám zdá, že je. Ilúzia je niečo, čo nie je tým, čím sa nám zdá že je“ (Blackmore, 2002). Z tohto dôvodu Blackmoreová odmieta teoretické predpoklady prívržencov súčasnej vedy o vedomí. Na rozdiel od Blackmoreovej, psychológ D. Wegner chápe zážitky vedomej vôľe ako pocity, ktoré nám pomáhajú zhodnotiť autorstvo toho, čo robíme. Pre Wegnera je dôležité porozumieť tomu, ako môže byť vedomá vôľa ilúziou, a nie to, že je ilúziou ako celok. Vedomé Ja je ilúziou v tom zmysle, že „skúsenosť s vedomým úmyslom vykonáť určitý čin nie je priamym náznakom toho, že tento čin spôsobil vedomá myšlienka“ (Wegner 2002, 2). Naznačená diskusia len v skratke odzrkadľuje pretrvávajúce nedorozumenia medzi teoretikmi a kulminovala až do absurdného vzájomného obviňovania sa z nedoceňovania či eliminácie fenoménu, ktorý sa stal legitímnym vedeckým explanandom. Zdrojom tohto, i v súčasnosti, neblahého stavu skúmania povahy vnútornej skúsenosti je podobne ako v dejinách filozofie zásadný rozpor medzi teoretickými východiskami filozofov, metódami a cieľmi skúmania. Vo vzťahu k riešeniu problematiky vedomej skúsenosti a Ja sa vo filozofii myšle udomácnili dva základné prístupy, takzvaná perspektíva prvej osoby (first person view) a perspektíva tretej osoby (third person view)². Perspektíva alebo pohľad prvej osoby sprevádzza človeka v každodennej skúsenosti pri uvedomovaní si toho, že je vlastníkom myšlienok, pocitov, predstáv, ktoré sú často neopakovateľné a niekedy pojмami neuchopiteľné.

² Schematickosť uvedenej dištinkcie nijak neproblematizuje používanie ďalších perspektív, ako napríklad perspektívy druhej osoby (second person view), ktorá tvorí jednu zo základných metód v psychiatrii a psychoterapii. Bližšie pozri in: Kircher et.al. (2003).

Perspektíva tretej osoby je súčasťou širšie uplatňovanej úspešnej stratégie empirickej vedy. Veda pristupuje k organizmom disponujúcim vedomou skúsenosťou vrátane Ja podobne ako k ostatným biologickým a fyzikálnym objektom sveta. Napriek svojej špecifickosti sa svet vnútornej skúsenosti pokladá za principiálne poznateľný, a teda podliehajúci metodológií bádania v súlade s ostatnými prirodzenými javmi. Rozdiely pri „čítaní“ povahy uvedenej dištinkcie sa premetli do sice dlhotrvajúcej, no nie veľmi plodnej diskusie. Podľa tých filozofov, ktorí uprednostnili epistemické dištinkcie prvá osoba - tretia osoba, existuje hlboká pripasť medzi poznávaním stavov myслe z perspektívy oboch pohl'adov. Názor, podľa ktorého je daná pripasť neprekonateľná priviedol niektorých filozofov k myšlienke kognitívnej uzavretosti (McGinn, 1993), t. j. principiálnej nemožnosti objasniť povahu stavov ľudského vedomia. Kameňom úrazu sa však podľa môjho názoru stalo metafyzické čítanie ústiacie do postulovania ontologickej pripasti medzi prvou osobou a tretou osobou. V dôsledku toho odlišnosti v poznávaní „nabudli“ status odlišnej existencie dvoch svojbytných svetov (svet prvej osoby a svet tretej osoby) alebo dát (dáta prvej osoby a dáta tretej osoby) (Searle, 1992). Rozkol v ontológii vyústil do radikálnych interpretácií filozofických konceptií. Stúpenci perspektívy tretej osoby sa kritizujú za to, že neberú vedomé Ja a „fenomenológiu vážne“ (Chalmers 1996), alebo za opomínanie „neredučovateľnosti príbehu prvej osoby“ (Velmans, 2004). M. Velmans, napríklad tvrdí, že D. Dennett vyhlásil prístup prvej osoby k fenomenálnym vlastnostiam za nevhodný pre vedu tretej osoby a teda za krajne nevhodný pre pochopenie vedomia vôbec. Pre fenomenológov je vedomá skúsenosť spolu s Ja ontologicky prvoradou vecou. Z tohto uhla pohl'adu sa potom zdá, akoby filozofi ako D. Dennett alebo P. Churchland (1995) „fenomenológiu neprikladali viero hodnosť“ či dokonca ju zo sveta eliminovali.

3 Mysel' v tele, telo v mysli a svet

Z naznačenej diskusie sa vykryštalizovali typy *Ja* označené pojмami, ktorých význam sa neraz prekrýva, ako napríklad: a) ekologické, b) privátne, c) senzomotorické, d) narratívne, e) fenomenálne, f) intersubjektívne, g) neurónové atď. *Privátne ja* patrí k najfrekventovanejším filozofickým pojmom, ktoré sa spája so schopnosťou človeka introspekтивne (vedome) vnímať alebo pozorovať obsahy vlastnej skúsenosti. Ako som uviedla v predchádzajúcej časti textu, v diskusiách sa často zmiešava jedinečnosť subjektívnej skúsenosti s bezprostredným a zároveň neomylným poznáním jej povahy. *Ekologické ja* sa vymedzuje na pozadí existencie aktívneho vzťahu medzi vnímajúcim a štruktúrovaným fyzikálnym a sociálnym prostredím. Výskum sa týka otázok o vzťahu medzi seba-

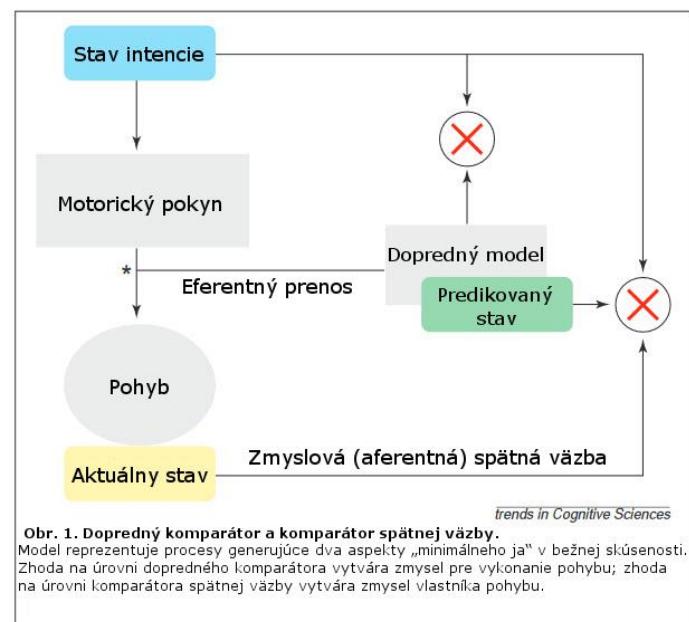
vnímaním a konceptom seba samého (self-concept) a takisto vývinom medzi perceptuálnymi a konceptuálnymi úrovňami seba-vedomovania. Americký psychiater D. Stern (1985) ho označil ako *základné ja*, ktoré tvoria tieto zložky: 1. pocit konajúceho (self-agency) v zmysle autorstva vlastných činov, 2. seba-koherencia (self-coherence) v zmysle existencie ohraničeného fyzikálneho celku, 3. seba-afektivita (self-affectivity) a 4. autobiografia (self-history)³. Aj psychológ U. Neisser používa vo svojej koncepcii pojem „ekologického ja“, pričom významné je to, že stavy a procesy, ktoré sa týmto pojmom zastrešujú majú pred-lingvistickú a nekonceptuálnu povahu. Podobne sa používa aj pojem *senzomotorického ja* podielajúceho sa na udržiavaní stability tela v priestore. Zvláštnu pozornosť venujú filozofi a literárni vedci pojmu *narratívneho Ja*. Vo všeobecnosti sa chápe v zmysle viac či menej koherentného Ja (seba-obrazu) zasadeného do času, do minulých a budúcich príbehov, ktoré o sebe a o druhých rozprávame. Vo filozofii myслe tento typ spopularizoval D. Dennett, ktorý chápe Ja v zmysle abstraktného „centra narratívnej tiaže“ kde sa stretávajú príbehy, ktoré rozpráva jedinec o sebe a o druhých. Prostredníctvom jazyka sme schopní vniest' relatívnu koherentnosť do našej skúsenosti. Rozprávaním príbehov rozširujeme svoje biologické hranice a vnášame do vlastného života zmysel. Súčasná literárna veda pracuje s pojmom narácie v rámci *kognitívnej narratológie* pri riešení problémov vzťahu narácie a identity, autobiografickej narácie, narratívnej konštrukcie reality atď. (Fonjoková, 2014). K naproblematickejším pojmom súvisiacim s vysvetľovaním povahy Ja patria tradične pojmy seba-vedomia (self-consciousness), seba-poznania (self-knowledge), seba-vedomenia (self-awareness), seba-kontroly (self-control) a neposlednom rade samotný pojem *selfa* a jeho vzťah k ostatným typom Ja.

Pri objasňovaní štruktúry, úrovní a typov Ja sa vynára potreba uplatňovať a spájať viaceré metódy a prístupy. Ako zdôrazňuje A. Damasio, súčasný výskum povahy vedomej skúsenosti a Ja vyžaduje integráciu viacerých perspektív, a to: 1) priameho vhl'adu individuálnej vedomej myслe, ktorá je privátна, osobná, jedinečná pre každého človeka (privátne ja), 2) behaviorálneho postoja, ktorý nám umožňuje pozorovať správanie druhých založené na predpoklade o existencii vedomých myслí podobných tej našej (intersubjektívne ja), 3) perspektívy mozgu, na základe ktorej môžeme skúmať aktivitu mozgu u indívduí, u ktorých predpokladáme prítomnosť alebo absenciu stavov myслe a vedomia (normálne, patologické ja) a 4) poznatkov evolučnej biológie a neurobiológie založených na skúmaní evolučnej histórie žívych organizmov, špecifík

³ Ekologický prístup priniesol nové fakty vo vzťahu k štúdiu a porozumeniu skúseností malých detí. Ako dokladá výskum, novorodenči sú už počas prvej hodiny schopní imitovať gestá tváří (neonatálna imitácia).

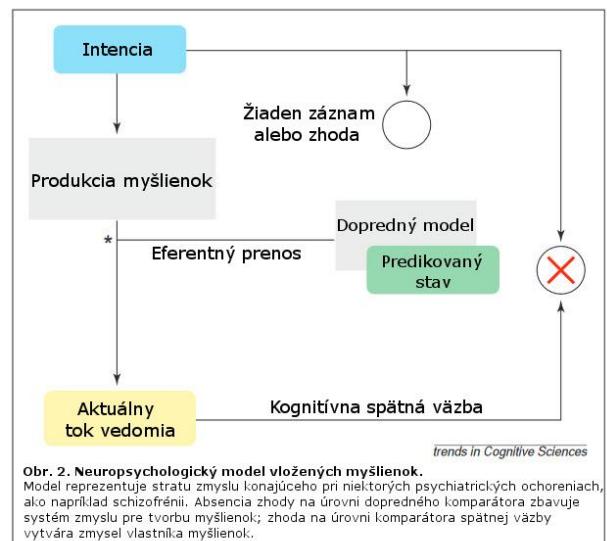
nervových systémov vo vzťahu k pôvodu myслe, ja a správaniu (základné, rozšírené ja). Vyústením uvedených perspektív je pracovná hypotéza, podľa ktorej sú mentálne udalosti ekvivalentné s príslušnými udalosťami v mozgu. Podľa Damasia možno povedať, že „niektoré neurónové vzorky sú simultánne mentálnymi obrazmi“ (Damasio, 1999, 16). Uplatnenie evolučnej perspektívy sa premetia do sofistikovanejších podôb, typov, úrovni a mechanizmov Ja. V naturalisticky ladených teoretických diskusiách vystupuje do popredia - dualita ja, t. j. rozlišovanie medzi dvomi evolučne vzniknutými Ja, a to: a) „ja ako objekt“ a b) „ja ako poznávajúci subjekt“⁴. Ja v zmysle poznávajúceho subjektu má svoj pôvod a teda vzniká v podobe novej neurónovej vrstvy ponad „ja ako objekt“. Heterogénnosť typov Ja sa premetla aj do úvah o podobe štruktúry jednotlivých stavov Ja. V úsilí sformulovať kritériá *minimálneho Ja*, S. Gallagher poukazuje na odlišnosť medzi dvomi úrovňami štruktúry ja – pocitom *konajúceho* (self-agency) a pocitom *vlastníka* (self-ownership). V prvom prípade jedinec pocítí, že on sám je tým, kto *zapríčinuje* alebo *vykonáva* aktivitu (napr. som to ja, kto zapríčinil pohyb lopty alebo som to ja, kto je pôvodcom vlastných myšlienok). V druhom prípade ide o pocítovanie toho, že ja som *vlastníkom* svojich predstáv, myšlienok a činov. V každodennej skúsenosti sa obe úrovne Ja vzájomne prelínajú (napr. uvedomujem si fakt, že moje myšlienky, rozhodnutia výstupili do istého typu správania, konania). K diskontinuite medzi pocitom *vykonávajúceho* a pocitom *vlastníka* môže dôjsť vtedy, ak pohyb sice pripríšeme sami sebe, teda ide o náš pohyb, ale výkonanie pohybu nemáme pod kontrolou (napr. pri lekárskom vyšetrení). Absencia pocitu *vlastníka* môže byť konzistentná s pocitom *konajúceho*. Ako však uvádzá Gallagher, v prípade takých javov ako sú sluchové halucinácie, vkladanie myšlienok sa objavujú problémy predovšetkým s pocitom *konajúceho*. Niektorí schizofrenickí pacienti trpiaci pocitmi, že im myšlienky „niekto“ *vkladá* (thought insertion), sú náchylní k omylom vo vzťahu k pocitom *vlastníctva*. Na ilustráciu Gallagher uvádzá neurokognitívny model bezprostredného seba-uvedomenia C. Fritha (Frith, 1992), ktorý prepája existenciu minimálneho seba-uvedomenia (self-awareness) s procesmi motorickej aktivity (obr.1).

⁴ V dejinách filozofie sa stretávame s rôznymi podobami duality Ja-objekt, Ja-subjekt v rámci fenomenologického, náboženského, morálneho alebo transcendentálneho ponímania Ja.



Obr.1 (Upravené podľa Gallagher, S., 2000)

Model vychádza z predpokladu, podľa ktorého možno pocíťovanie vlastníctva pri motorickej aktivite vysvetliť v pojoch ekologického seba-uvedomenia zabudovaného do pohybu a vnímania. Z experimentálneho výskumu normálnych subjektov sa ukazuje, že pocit *vlastníka* pri konaní vychádza z toho, čo predchádza konaniu a premieňa intenciu na aktivitu. Na základe uvedeného, ako konštatuje Frith, schizofrenickí pacienti trpiaci pocitom *vložených myšlienok*, sa dopúšťajú omylov aj v súvislosti s výkonávaním rôznych telesných pohybov, majú problémy s dopredným monitorovaním pohybu predchádzajúcemu konaniu, ale nie s motorickou kontrolou založenou na porovnávaní zámerného pohybu a zmyslovej spätnej väzby (obr.2).



Obr.2 (Upravené podľa Gallagher, S., 2000)

Uvedený model, podľa Gallaghera, podporuje vymedzenie *minimálneho ja* ako skúsenosti prežívania subjektu v zmysle bezprostredného vzťahu k sebe samému, nerozpriestraneného v čase. Ja v zmysle statického objektu alebo veci sa napokon nahrádza obrazom dynamického komplexného procesu prebiehajúceho na viacerých úrovniach skúsenosti. Táto skutočnosť podľa môjho názoru výrazne prispieva k vysvetľovaniu povahy Ja v tom, že o *skúsenosti-seba* má zmysel hovoriť aj vtedy, keď si subjekt vlastnú skúsenosť neuvedomuje.

4 Ja a vedomie

Americký neurológ A. Damasio vo svojom prístupe k fenoménu Ja prepojil poznatky z neurologickej praxe so sofistikovaným pohľadom na vznik, mechanizmy a funkcie vedomia. Pre Damasia je vedomie synonymické s pociťovaním toho, čo sa stane, s poznaním toho, že máme emócie, že sa radujeme, trápime a konáme. Zrod poznávajúceho Ja (vedomia) prirovnáva k hercovi, ktorý vystúpi na scénu spoza opony, z tmy na svetlo. V súvislosti so skúmaním vedomia sa vynárajú dva základné problémy: 1. porozumieť tomu, ako mozog generuje mentálne vzorky – obrazy (images) objektu, osoby, ako vytvára neuronálne vzorky v neuronálnych okruhoch a ako ich mení na explicitné mentálne vzorky a 2. akým spôsobom mozog generuje pocit Ja v procese poznávania. Vo svojej teórii popisuje vytváranie „predstavenia v mozgu“ a vznik „jeho vlastníka a pozorovateľa“ (Damasio, 1999, 11). Evolučná perspektíva, z ktorej vo svojom prístupe vychádza viedie Damasia k rozlíšeniu viacerých typov vedomia, a to: a) základné vedomie (core consciousness) a b) rozšírené vedomie (extended consciousness). Nositel' evolučne staršieho základného vedomia prežíva svoj život „tu a teraz“, t. j. v prítomnosti bez zamýšľania sa nad budúnosťou. Základné vedomie je jednoduchý biologický jav s jedinou úrovňou organizácie, je stabilný počas života a neprináší výlučne človeku. Základné vedomie nepotrebuje jazyk, pamäť, pozornosť či rozum. Rozšírené vedomie má viacero stupňov a úrovni kulminujúcich v schopnosti organizmu anticipovať budúcnosť a v identite osoby. Evolučne vyšší typ vedomia sice závisí od nižšieho typu, no jeho narušenie ho ponecháva nedotknuté. Opak však neplatí, poškodenie základného vedomia prináša súčasne poškodenie rozšíreného vedomia. Príkladom poškodenia rozšíreného vedomia pri zachovaní základného vedomia je napríklad anosognózia (z gréck. *nosos*, choroba a *gnosis* poznanie) – nerozpoznanie choroby vo vlastnom tele.

U pacienta po porážke (poškodenie pravej hemisféry) dochádza k paralýze ľavej strany tela. Pacient nie je schopný hýbať nohou, rukou, polovicou tváre, chodiť ani stáť, ale o svojom stave hovorí, akoby sa nič neprihodilo. Damasio

interpretuje výpoved' pacienta ako stratu kognitívnej funkcie v dôsledku poškodenia špecifickej oblasti mozgu. K ďalším poškodeniam rozšíreného vedomia sa radia niektoré typy mánie, depresie alebo schizofrénie (napríklad depersonalizácia, poškodenie Ja). Pri zmenách rozšíreného vedomia zostáva zachovaná bdelosť, nízkoúrovňová pozornosť, tzv. background emócie (stav pokoa alebo napätia) a rutinné správanie. Problém nastáva pri vysoko špecializovanom správaní závislom od poznania minulosti a budúcnosti. Mentálna reprezentácia autobiografického ja je tak narušená, že „mysel' nevie odkiaľ toto ja pochádza ani kam smeruje“ (Damasio, 1999, 217). Za základnú podmienku prítomnosti vedomia v organizme pokladá Damasio „schopnosť ukazovať si v mozgu „obrazy“ a objednávať si ich v procese myslenia“ (Damasio, 2012). Obrazy sú v skutočnosti usporiadane „vzory“ (patterns) činnosti miliónov synapsí na viacerých vzdialených miestach mozgu. S relatívne simultánou aktivitou prebiehajúcou v rôznych oblastiach mozgu sú prepojené jednotlivé stavy mysle. Napríklad, ak sa má organizmus rozhodnúť ako konáť, musí disponovať rozsiahlymi znalosťami o sebe, teda o vnútornom organizme ako aj o vonkajšom svete. V záujme prežiť zvíťazí také usporiadanie mozgu, v ktorom subsystémy zodpovedné za uvažovanie a rozhodovanie budú tesne prepojené so subsystémami, ktoré rozhodujú o biologickej regulácii. Význam Damsiovej teórie vedomia spočíva v prepojení vzniku a vývinu vedomia so vznikom a vývinom Ja . K základným typom Ja zaraďuje: 1. proto-ja (proto-self), 2. základné ja (core-self) a 3. autobiografické ja (autobiographical-self). Posledným dvom typom ja zodpovedajú aj dva typy vedomia, základné vedomie a rozšírené vedomie. 1. Predvedomým biologickým predchodom Ja, ktoré vo všeobecnosti pokladáme za špecifikum ľudského vedomia, sa stáva neuronálne ja (proto-ja). Vytvára ho súbor neuronálnych vzoriek mapujúcich aktuálny stav fyzikálnej štruktúry organizmu, odohráva sa na rozmanitých úrovniach v mozgu - od mozgového kmeňa k cerebrálnemu kortextu, v štruktúrach prepájaných neuronálnymi dráhami. Proto-ja si neuvedomujeme, jazyk nie je súčasťou jeho štruktúry, podiel'a sa na vzniku reprezentácie stavov tela prvého rádu, ide o stavy, pri ktorých sa čosi poznáva. Vedomými sa podľa Damasia stávame vtedy, keď vykazujeme špecifický druh poznania. Kontakt s objektom spôsobuje zmeny v organizme. Reprezentácie objektu vytvárajú jednu mapu – prvej úrovne a ponad ňu a reprezentácie organizmu druhú mapu. Proto-ja sa mení vtedy, keď vzniká reprezentácia vzťahu objektu a proto-ja v čase, dochádza k modifikácii proto-ja objektom. Vzniká základné ja (core self), ktorému zodpovedá základné vedomie, t.j. neverbálny prístup druhej úrovne objavujúci sa vždy pri uvedenej modifikácii proto-ja. Mechanizmus produkujúci základné ve-

domie prechádza počas života minimálnymi zmenami. Základné ja si uvedomujeme. Je kontinuálne v čase. Mechanizmus základného vedomia zahŕňa prítomnosť proto-ja. Biologickou podstatou sú reprezentácie máp druhej úrovne. Vo vývinovom reťazci vzniká ďalej autobiografické ja, ktorému zodpovedá rozšírené vedomie. Autobiografické ja je založené na autobiografickej pamäti, zahŕňa individuálnu skúsenosť minulého a anticipáciu budúcnosti, na permanentných no dispozičných záznamoch skúseností základného ja. Vyžaduje prítomnosť základného ja a jeho mechanizmov. Za jednu z najvýznamnejších vlastností neurónovej aktivity mozgu pokladá Damasio schopnosť vytvárať mapy (Damasio, 2012, 63-67). Mozog tak plní funkciu kartografa, ktorého rola spočíva v informovaní seba samého, pričom základ všetkých informácií sa spracúva na nevedomej úrovni.

5 Patologické ja

Už v predchádzajúcim texte som uviedla niektoré poškodenia súvisiace s našou vnútornou skúsenosťou Ja. Je pozoruhodné, že nielen súčasní filozofi a kognitívni vedci, ale aj psychiatri ako napríklad G. Berrios a I. Marková (Kircher, 2003) nepokladajú pojem „Ja“ za prirodzený druh (natural kind), ale skôr za lingvistický tróp. Metaforické chápania Ja počas reformácie v 17. storočí, používanie Ja v sociálno-politickej kontexte 18. storočia sa prenesli do psychiatrie v 19. storočí v zmysle psychologickej entity a začal sa proces jeho „naturalizácie“. Domnenka o existencii Ja vo vnútri myseľ a mozgu sa premietla do chápania Ja ako entity, ktorú možno ovplyvňovať patologickými léziami a chorobami. Patológia Ja sa v medzinárodnej klasifikácii duševných chorôb (MKCH-10-SK-2016) radí k poškodeniam súvisiacim s poruchami osobnosti. Podobne ako v rámci rozmanitej typológie poškodení osobnosti, poškodenia Ja sa vyskytujú na jednotlivých úrovniach štruktúry Ja, o ktorých som sa zmienila v predchádzajúcim teste. O poškodení naratívneho Ja svedčí syndróm „rozštiepeného mozgu“ (split-brain). Keďže syndróm zahŕňa rečové centrum ľavej hemisféry, M. Gazzaniga ho označil pojmom „interpretátor“. U pacientov, ktorí majú oddelené hemisféry mozgu, ľavá hemisféra vystupujúca ako „sídlo“ naratívneho Ja je schopná udržať ilúziu vlastnej stability (Gazzaniga, 2011), resp. koherenciu a stabilitu vlastnej vedomej skúsenosti. K neurologickým prípadom, pri ktorých je toto úsilie narušené patria syndrómy „cudzej ruky“ (alien-hand) - pacient nepokladá vlastnú končatinu za svoju alebo „strata ja“ (loss of self) – strata zmyslu kym sme, kto sme – sprevádzané pocitmi seba-odcudzenia, zmeny veľkosti končatín, vnímanie vlastnej osoby zdial'ky, neschopnosť kontrolovať vlastnú aktivitu. Diagnostika viacerých poškodení Ja dokladá význam kontextu skúsenosti, t. j. toho, ktoré faktory pri-

narušení kontextu môžu viest' k vzniku poškodenia Ja. Napríklad k depersonalizácii dochádza narušením kontextu spôsobeným zlyhaním očakávaní, depresiou alebo bolest'ou a pod. Objavuje sa frekventované u starších adolescentov, ktorí prechádzajú obdobím utvárania ciel'ov a zmyslu svojho ďalšieho života. Aj z vlastnej skúsenosti vieme posúdiť dôsledky neopäťovanej lásky, ktoré môžu vyústiť až do „rozpadu sveta“ odmiennutej osoby. Pri fúge (fugue) dochádza k neočakávanému výpadku pamäti súvisiacej s identitou osoby, s neschopnosťou spomenúť si „kto som“. Častokrát sa objavuje spolu s psychologickou amnéziu po silných stresoch, prírodných katastrofách a pod. Viacnásobné ja (multiple selves) sa prejavuje pociťovaním existencie viacerých osobností ako integrovaných, komplexných „jednotiek“ s premenlivou dominanciou v rámci jedného indívuda (napríklad vyvolané traumatickými skúsenosťami z detstva). Ukazuje sa, že rôzne úrovne a typy ja spolupracujú, naratívne ja môže napríklad kompenzovať poškodenia senzomotorického ja. U pacientov s neglektom ľavého osobného priestoru, napríklad pri paralíze ľavej strany tela v dôsledku mŕtvice v pravej hemisfére, pacienti neraz trvajú na nesprávnej interpretácii svojej paralízy, napr. vlastnú končatinu pokladajú za cudziu alebo prináležiacu inej osobe. Anosognózia sa charakterizuje ako poškodenie Ja charakterizované stratou vhl'adu vyšej rovne (higher-order insight) na jestvujúci deficit. Keďže poškodenie sa týka osoby samej vymedzuje sa nedostatkom informácie na úrovni sebareprezentácie (pozornosti, kognitívnej prístupnosti). V prípade Antonovho syndrómu dochádza k popieraniu slepoty. Pacient v dôsledku kortikálnej slepoty a následným bilaterálnym poškodením okcipitálnych lalokov sa stáva neschopným spracovať vizuálnu informáciu a správa sa podobne ako pacient s periferálnou slepotou. Pacienti vykazujúci známky slepoty, narážajú do nábytku, niektorí pacienti svoju slepotu popierajú a správajú sa tak akoby dané poškodenie neexistovalo. Dochádza ku konfabulácii zrakových skúseností, akoby „vymiznutie“ vizuálneho sveta nebolo niečim, čoho nie sú si vedomí, zároveň vypovedajú o neexistujúcich fenomenálnych svetoch, v existenciu ktorých, ako sa zdá, veria (Goldberg, 2009). Experimentálny výskum podporuje myšlienku asymetrie medzi introspektívne prežitými vedomými stavmi a teoretičky založenými vysvetleniami pretrvávajúcich príčinných procesov. Ako dokladá neurokognitívny výskum, prežitá priorita myšlienok pred činom, konzistencia myšlienok o plánovanom čine obsahuje skôr magické než príčinne súvislosti pri vysvetľovaní správania človeka. D. Wegner to výstižne označil takto: „Skúsenosť vôle je ako mágia“ (Wegner, 2002, 289). Štúdium normálneho a poškodeného vedomia naznačuje, že zážitok vedomej, zámernej aktivity sa môže objaviť: a) pred činom, b) hned' po čine, c) počas činu. Osoba môže poci-

ťovať vedomú vôľu v číne, ktorý nepredvídala (konfabulácia zámerov), alebo necíti zodpovednosť za vykonané činy - „syndróm cudzej ruky“ (alien-hand syndrom). Ukazuje sa, že potreba človeka zdôvodniť svoje postoje a konanie je v skúsenosti bežne prítomná. Vedci a teoretici na príklade viacerých experimentov predviedli spôsoby, akými možno oddeliť činy od prežívania alebo vôle. Ľudia pociťujú vedomú vôľu v činoch, ktoré nepredvídali, napríklad, pri konfabulácii zámerov sa snažia pozmeniť, čo mali v úmysle vykonať po tom, čo dokončia svoj čin. Pri „posthypnotickej sugescii“ hypnotizovaný subjekt súhlasí s pokynmi a často prichádza s vlastnými vysvetleniami činov nasledujúcich po sugescii. Pacient so „syndrómom cudzej ruky“ (Sacks, 1985) v činoch necíti vedomú vôľu, svoju ruku cíti ako cudziu a vykonáva s ňou činy, za ktoré je zodpovedný, ale neuvedomuje si to. Pri „syndróme uzamknutia“ je pacient pri vedomí, má v úmysle konáť, ale nie je toho schopný a trpí (Damasio, 1999). Každodenné skúsenosti prinášajú množstvo príkladov nevedome vykonávaných činov, pri ktorých ľudia prehodnocujú svoje zámery, aby tak odôvodnili svoje konanie. Ak osoba vykoná niečo, čo je v rozpore s jej predchádzajúcim zámerom, je znepokojená a snaží sa vymníť myšlienkom v rozpornom vzťahu (kognitívna disonancia). Klasické prípady nevedome vykonávaných činov boli skúmané v automatizmoch, pri ktorých absencia prežitého zámerného činu mala za následok pripisovanie príčin a konania pri udalostach tajomným silám (počas spirituálnych seáns, automatického písania, atď.). Štúdium a vysvetlovanie „ideomotorických činov“ spochybnilo nadprirodzené interpretácie a otvorilo cestu vedeckému skúmania týchto javov. („Špión“ teda odhalil ilúzoru podstatu duševného príčinného reťazca – ilúziu príčinného/časového sledu, takpovediac, zlava doprava - od myšle k správaniu. Ako bolo predvedené, skúsenosť vedomého úmyslu pre čin a kauzalita tohto činu vyvolaná vedomou myšľou osoby nie sú rovnaké procesy. Tendencia zamieňať si ich je podľa Wegnera zdrojom ilúzie vedomej vôľe. Zážitok vedomej vôľe je ilúziou v tom zmysle, že je vytvorený vnímaním zdanlivých príčinných procesov prepájajúcich vedomé myšlienky osoby s jej činmi. Ako tvrdí Wegner, v skutočnosti vôbec nemusí ísť o príčinný mechanizmus. Disponujeme sice myšlienkami, rozhodnutiami o tom, čo budeme robiť a zväčša ich aj pokladáme za stavy s príčinnou silou, pritom však môže ísť o predbežnú možnosť konania, ktoré nasleduje. To, že pocit vôľe nie je priamo prepojený s príčinným mechanizmom, nám pripomína Dennettov (Dennett, 1991) koncept zaujatia „intencionálneho postoja“ k ľuďom, na základe ktorého ich pokladáme za agensov s túžbami a úmyslami zapričinujúcimi konanie.

Tradičné teórie myšle vychádzajú z konceptu „epistemickej transparentnosť“ pri skúmaní seba-vedomia (uvedomovania), t. j. predpokladu, podľa

ktorého vnútorné ja (prežívanie ja v skúsenosti) presvetľuje skrz-naskrz svetlo poznania. V dejinách filozofie, ale aj psychológie sa metafora svetla udomácnila pri spájaní svetla so slobodou, s poznáním a seba-poznaním. Ako som však uviedla na viacerých miestach v texte, niektoré z poškodení Ja spochybňujú explanačnú silu takto poňatého vzťahu myšle a transparentnosti. Domnele pozitívna hodnota osvetenia pri pohľade do hlbín nášho vnútra môže zavádzat, zaslepiať. Aj preto T. Metzinger hovorí o transparentnosti ako o temnote a navrhuje rozlíšiť viaceré typy transparentnosti, a to minimálne epistemickú a fenomenálnu transparentnosť. Epistemická transparentnosť sa viaže na schopnosť pozorovať alebo vnímať obsahy vlastnej skúsenosti. Teoreticky sa daná schopnosť označuje pojmom introspekcie. Fenomenálnu transparentnosť charakterizuje bezprostredný prístup ku všetkým stavom, ktoré prežívame, či ide o skutočné objekty v okolitej svete alebo obrazy rozprávkových bytosťov objavujúcich sa v našich snoch. V tomto svete nemá zmysel rozlišovať medzi javením a realitou, nakoľko nejde o prístup, prostredníctvom ktorého odhalujeme povahu veci samej. Aj vtedy, a možno najmä v prípade, ak je skúmaný fenoménom skúsenosť, skúsenosť Ja sama. Zamieňanie epistemickej s fenomenálnou transparentnosťou má za následok zmiešavanie poznávacej, „auto-epistemickej“ uzavretosti s aktuálne prežívanou fenomenálnou skúsenosťou. Transparentnosť v zmysle temnoty vychádza na povrch predovšetkým vtedy, keď sme presvedčení o neomylnosti vlastných počítov či úsudkov, v prípadoch keď si neuvedomujeme (alebo si nechceme uvedomiť) omylnosť obšahov vlastnej myšle.

6 Zhrnutie

Súbežne s nárastom a prehlbovaním typológie Ja sa vykryštalizovali viaceré úrovne alebo štádiá vo vývine Ja. Ukazuje sa, že Ja predstavuje komplexnú štruktúru, ktorá koreluje s mnohovrstvenatou neurónovou aktivitou a je podmienená celým radom faktorov. Napriek počiatočnému stavu vedeckého a teoretického skúmania povahy Ja možno už v súčasnosti vypozorovať hlavné línie a prístupy výskumu. Najvýznamnejšie poznatky poukazujú na komplexnosť otázok vzniku, štruktúry, typológie, vývinu a funkcií fenoménu Ja. Experimentálny a klinický výskum normálneho a patologického Ja je dobrým dôvodom očakávania postupného pokroku vo vedeckom, ale aj teoretickom bádaní. To predpokladá permanentné odbúravanie zavádzajúcich intuícií pri vymedzovaní kritérií na určenie minimálneho Ja a vedomia, javov, ktoré sa neustále vyvíjajú a menia. K frekventovaným omylem patrí napríklad: a) identifikácia subjektivity s vedomou skúsenosťou, b) nerozlišovanie medzi epistemickým a fenomenálnym Ja, c) prehliadanie funkcií

a roly nevedomých zložiek a procesov v kontexte vnútornej skúsenosti človeka.

Ak sa usilujeme porozumieť hlbšie vedomej skúsenosti človeka, uvedomovaniu si vlastného Ja, potreba permanentnej konfrontácie poznatkov kognitívnovedného, klinického výskumu s filozofickými prístupmi a modelmi je evidentná. Za predpokladu, samozrejme, že sa nebudeme brániť radikálnemu prehodnoteniu „ideálneho modelu“ človeka ako racionálne uvažujúceho, vedomého, zodpovedného a slobodného agensa.

Literatúra

Blackmore, S. (2002). There Is No Stream of Consciousness. What is all this? What is all this stuff around me; this stream of experiences that I seem to be having all the time? *Journal of Consciousness Studies*, 9, Numbers 5-6, 1, pp. 17-28(12).

Blanke, O. (2012). “Multisensory brain mechanisms of bodily self consciousness”. *Nature Reviews Neuroscience*. 13: 556-571.

Chalmers, D. (1995). “Facing up to the problem of consciousness”. *Journal of Consciousness Studies*, 2, 3: 200-219.

Damasio, A. (1999). *The Feeling of What Happens*, London, William Heinemann.

Damasio, A. (2012). *Self Comes to Mind: Constructing the Conscious Brain*. New York, Vintage Books.

Dennett, D. (1991). *Consciousness Explained*. Boston, Little, Brown and Company.

Fojnoková, Z. (2014). Narativní konstrukce identity. In: Kuzmíková, J. (ed.) *Literatúra v kognitívnych súvislostiach*. Bratislava, Ústav Slovenskej literatúry SAV, s.74-100.

Frith, C. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Erlbaum.

Gallagher, S., Zahavi, D. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London, Routledge.

Gallagher, S. (2000). Philosophical Conceptions of the Self: Implications for Cognitive Science. *Trends in Cognitive Sciences*. 4(1): 14-21.

Gazzaniga, M. (2011). *Who's in Charge?: Free Will and the Science of the Brain*. New York, NY: Ecco.

Goldberg, E. (2009). *The New Executive Brain: Frontal Lobes in a Complex World*. New York Oxford University Press.

Kahneman, D., Slovic, P., Tversky, A. (1982): *Judgment under uncertainty: Heuristics and biases*. Cambridge, Cambridge University Press.

Lakoff, G., Johnson, M. (1999). *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*. New York, Basic Books.

McGinn, C. (1991). *The Problem of Consciousness*. Basil Blackwell.

Metzinger, T. (2010). *The Ego Tunnel*, Basic Books, New York.

Sacks, O. (1985). *The Man Who Mistook His Wife for a Hat*. Summit Books.

Stern, D. (1985). *The Interpersonal World of the Infant*. New York, Basic Books.

Strawson, G. (2010). “Radical Self -Awareness”. In: S. Mark, E. Thompson, D. Zahavi (eds.) *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford, Oxford University Press, 274-307.

Varela, F. (1996). “Neurophenomenology. A Methodological Remedy for the Hard Problem”. *Journal of Consciousness Studies*, 4(3): 330-349.

Varela, F., Thompson, E., Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge MA, The MIT Press.

Velmans, M. (2009). “How to define consciousness – and how not to define consciousness”. *Journal of Consciousness Studies*. 16(5): 139-156.

Velmans, M. (2004). “Why conscious free will both is and isn’t an illusion”. *Behavioral and Brain Sciences*, 27(5): 649-659.

Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MIT Press.

Wilkes, K. (1988). *Real People: Personal Identity Without Thought Experiments*. Oxford, Oxford University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford, Blackwell.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA, MIT Press.

Improving the performance of impulse neuro–glial network

Peter Gergel[†]

Department of Applied Informatics, Comenius University in Bratislava

Mlynská dolina, 84248 Bratislava, Slovak Republic

Email: peter.gergel@gmail.com

Abstract

Research in neuroscience over the past few decades has shed new light on glial cells which were always considered as purely passive supportive cells. New data provides evidence that astrocytes, a group of glial cells, possess important physiological functions that distinguish them from passive cells. It is now known that astrocytes are actively involved in neuronal communication regulation and synaptic transmission. Similar to neurons, astrocytes are integrated into networks that enable them to communicate with one another over long distances using Ca^{2+} signals. Since this is a relatively new area of research in neuroscience, computational models (mainly connectionist) are still missing. In this paper we briefly introduce astrocytes, focus on existing connectionist models and present technique that yields more successful performance.

1 Introduction

According to current knowledge of neuroscience, two types of brain cells are distinguished: neural cells and glial cells. Population of neurons is characterized by the ability to generate action potential whereas glia were always considered as passive non-excitatory cells. Neurophysiological findings at the turn of the 90s began to shift this perspective by providing evidence that glia cannot be longer considered as passive cells since they possess mechanisms for neural and synaptic regulation and modulation.

It is well-known notion that glia outnumber neurons by tenfold as taught by many neuroscience textbooks, but more recent evidence debunks this concept and claims glia-to-neuron ratio is in reality close to one (Hilgetag and Barbas, 2009). Four types of glial cells are distinguished: *oligodendrocytes*, *microglia*, *ependymal cells* and *astrocytes*. First three types are highly specialized and responsible for myelination, immunity and production of cerebrospinal fluid. Astrocytes account for most potent and functionally richest type that modulates neuronal activity on multiple levels. Besides their crucial role in neural development and production of neurotrophic factors, they are essential in regulation of extracellular level of pH and K^+ , glutamate metabolism, nurturing neurons and maintaining blood–brain barrier.

Important milestones in history of glial research included discovery of negative resting membrane potential in astrocytes and membrane depolarization by neural activity (Orkand et al., 1966) and formation of glial syncytium (Brightman and Reese, 1969). In terms of current knowledge astrocytes are considered to be chemically excitable by specific neural molecules to which they respond with increased intracellular Ca^{2+} oscillations and release of *gliotransmitters* that regulate and modulate neuronal and synaptic activity.

Despite the fact that glial research is for the last decades very popular in neuroscience, in the area of computational modeling lack of interest still persists. In this paper we focus on existing connectionist models of multi-layer perceptron (MLP) with *artificial astrocytes* and present technique that yields more successful performance.

2 Previous work

In the area of computational neuroscience two modeling paradigms (abstractions) are considered: *biophysical* and *connectionist*. While first paradigm focuses on physicochemical properties of biological system using mathematical formalizations, second one makes abstraction over low-level mechanisms and tries to comprehend the system from higher level.

Despite plethora of biophysical models of astrocytes and their interactions with neurons, connectionist models are still missing. For an overview of biophysical models I suggest Oschmann et al. (2017), Volman et al. (2012), Wade et al. (2014).

Although not considered as a computational modeling, some authors focus on modeling neural mechanisms using electronic circuits. Joshi et al. (2011), Irizarry-Valle et al. (2013), Irizarry-Valle and Parker (2015) present neuromorphic engineering and designed CMOS circuits to model small neural networks extended with astrocytes that modulate excitatory postsynaptic potential.

2.1 Multi-Layer Perceptron with Chaos Glial Network

Ikuta et al. (2010) proposed a concept of artificial astrocytes in artificial neural networks and have written multiple papers regarding this topic. Their proposed idea was to extend MLP with astrocytes on a hidden layer (Fig. 1) using formula 1 for computation of hidden neuron's output.

$$h_i(t+1) = f\left(\sum_{j=0}^n w_{ij}(t)x_j(t) + \alpha\Psi_i(t)\right) \quad (1)$$

Activation of hidden layer is determined by linear combination of weights and input from a previous layer summed with an astrocytic output, $\Psi_i(t)$, multiplied by weight α (shared amongst all astrocytes). Output for a single astrocyte is given by formulas 2 and 3.

$$\Psi_i(t) = \sum_{k=-m}^m \beta^{|k|} \psi_{i+k}(t) \quad (2)$$

$$\psi_i(t+1) = \begin{cases} \frac{2\psi(t)+1-A}{1+A}, & (-1 \leq \psi(t) \leq A) \\ \frac{-2\psi(t)+1+A}{1-A}, & (A < \psi(t) \leq 1) \end{cases} \quad (3)$$

where activation of i -th astrocyte, $\Psi_i(t)$, is a sum of m neighboring astrocytes, $\psi_i(t)$, with exponentially attenuating weight by factor $0 < \beta < 1$. Biological motivation using these formulas is that it is known astrocytes generate spontaneous Ca^{2+} oscillations (formula 3) that are propagated throughout glial syncytium (formula 2).

To assess performance of proposed model authors chose well-known benchmark *two-spiral problem* depicted by Fig. 2. This problem cannot be easily solved

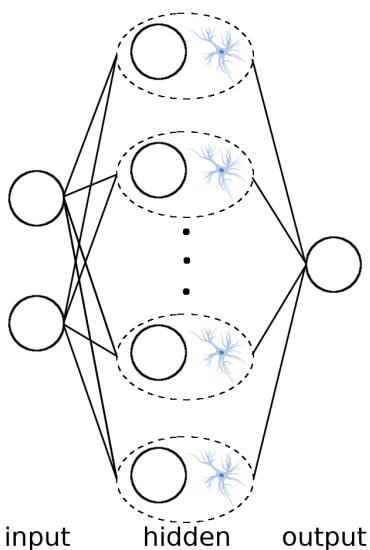


Fig. 1: Architecture of the proposed model. MLP with artificial astrocytes on a hidden layer.

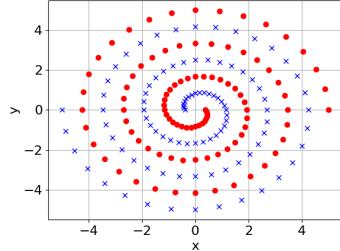


Fig. 2: Two-spiral problem: assigning every point (x, y) to class $\{0, 1\}$.

by standard MLP due to large number of local minima. Architecture of the proposed MLP consisted of two hidden layers with structure 2–20–40–1 with astrocytes extending the second layer exclusively using parameters $\beta = 0.8$ and $m = 5$. The model was trained using standard gradient descent algorithm – *backpropagation*.

The authors compared performance of conventional MLP, MLP with *random noise* and proposed MLP (with *chaotic noise*) using *mean squared error*. Proposed model yielded best results followed by MLP with random noise and conventional MLP.

2.2 Multi-Layer Perceptron with Impulse Glial Network

Ikuta et al. (2011) continued in the concept of artificial astrocytes, but instead of implementing chaotic oscillations they introduced *active* astrocytes that listen to and regulate neural activity. Formula 3 was substituted with formula 4.

$$\psi_i(t+1) = \begin{cases} 1, & \theta_n < h_i(t) \wedge \theta_g > \psi_i(t) \\ \gamma\psi_i(t), & \text{otherwise} \end{cases} \quad (4)$$

with 3 new parameters introduced: θ_n is astrocyte's activation threshold, θ_g is astrocyte's refractory period and γ is attenuation factor. Astrocyte generates 1 only when has recovered from refractory period and neuron's output overpasses manually chosen constant threshold.

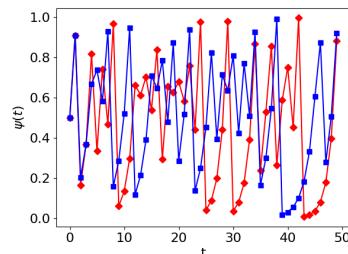


Fig. 3: Two time series used for evaluating performance of model. Task is to assign 4 adjacent values $(\psi(t), \psi(t+1), \psi(t+2), \psi(t+3))$ to classes $\{0, 1\}$.

The problem used for evaluating performance of this model was classification of 4 adjacent values into 2 time series (Fig. 3). The architecture of all 4 models was 4-10-1 and they were trained using backpropagation algorithm. Using metric MSE for estimating classification errors yielded in results displayed in Tab. 1.

Model	avg. MSE
Conventional	0.0351
Random Noise	0.0111
Chaotic Noise	0.0074
Impulse Glia	0.0053

Tab. 1: Performance of 4 models in classification task on time series. Conventional MLP yielded worst results and proposed model with impulse glia yielded best results.

2.3 Astrocyte-driven synaptic plasticity

Concept of synaptic plasticity driven by artificial astrocytes was researched by Alvarellos-González et al. (2012). They extended every neuron (including input and output layers) with astrocyte controlling its synaptic plasticity by counting how many times neuron fired for last k times:

$$r_j(t) = \sum_{i=0}^{k-1} u(x_j(t-i))$$

where $x_j(t)$ is an output of j -th neuron and $u(x_j(t))$ is defined as:

$$u(x_j(t)) = \begin{cases} -1, & x_j(t) \leq 0 \\ 1, & x_j(t) > 0 \end{cases}$$

Synaptic weights are changed accordingly to rules:

$$w_i(t + \Delta t) = w_i(t) + \Delta w_i(t)$$

$$\Delta w_i(t) = |w_i(t)|z(t)$$

$$z(t) = \begin{cases} a, & r_j(t) = \mu \\ b, & r_j(t) = -\mu \end{cases}$$

Parameters a , b , and μ were heuristically chosen. Authors proposed several rules of synaptic plasticity based upon this idea, but we are not listing them all.

Authors evaluated performance of the model trained using proposed rules on 4 different problems and confirmed that model with astrocytes (NGN) was able to learn problems more accurately than identical model without astrocytes (NN), but it highly depended on complexity of selected problem. In 3 cases (amongst 4) NGN was superior to NN. It is questionable why authors did not choose backpropagation algorithm for training NN model, but decided for genetic algorithm which is known to be less effective in training neural networks.

Since algorithm requires parameters a , b , and μ that need to be chosen manually for every individual problem, Mesejo et al. (2015) proposed technique based on evolutionary algorithms and implemented proof of concept that automatically searches for optimal values of parameters.

3 Optimizing astrocyte weight

In our research we were inspired by work Ikuta et al. (2010) described in Section 2.2. We have chosen MLP extended by impulse astrocytes at hidden layer. To assess performance of our model we focused on two-spiral problem (Fig. 2) using architecture 2–20–1. For computation of hidden layer’s output we used slightly modified version of formula 1:

$$h_i(t+1) = f(\sum_{j=0}^n w_{ij}(t)x_j(t) + \alpha_i\Psi_i(t))$$

Notice α was substituted by α_i . For activation of astrocytes we used formulas 2 and 4. Parameter values were: $\eta = 0.01$, $m = 20$, $\beta = 0.5$, $\theta_n = 0.6$, $\theta_g = 10^{-4}$, $\gamma = 0.5$. Since loss function is defined as:

$$E(w) = \frac{1}{2} \sum_p (d^{(p)} - y(x^{(p)}))^2$$

parameter α_i can be optimized:

$$\Delta\alpha_i = -\eta \frac{\partial E(w)}{\partial \alpha_i}$$

$$\Delta\alpha_i = -\eta \left(\sum_j \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial h_i} \right) \frac{\partial h_i}{\partial net_i} \frac{\partial net_i}{\partial \alpha_i}$$

$$\Delta\alpha_i = -\eta \left(\sum_j \overbrace{(t_j - y_j)y_j(1 - y_j)}^{\delta_j} w_{ji} \right) h_i(1 - h_i) \Psi_i$$

$$\Delta\alpha_i = -\eta \overbrace{\left(\sum_j \delta_j w_{ji} \right)}^{\delta_i} h_i(1 - h_i) \Psi_i$$

$$\Delta\alpha_i = -\eta \delta_i \Psi_i$$

We implemented proof of concept algorithm to computationally evaluate performance with results illustrated in Fig. 4.

4 Conclusion

Understanding the role of astrocytes and neuron–astrocyte interactions are essential for understanding the brain. Computational models in this context are inevitable since they allow us to test correctness of our knowledge, specified hypotheses and propose means for prediction of certain behavior. In this work we focused

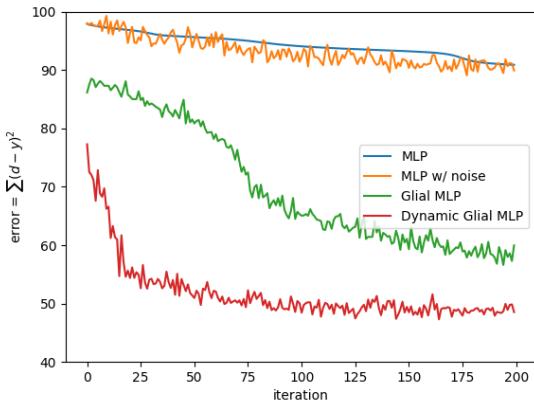


Fig. 4: Simulations of 4 models on two-spiral problem. Conventional model (blue) and model with gaussian noise (orange) were not able to escape local minima, but models with artifical astrocytes (green and red) were. Dynamic update of α_i during training (red) yielded lowest error (best performance).

on existing connectionist models that incorporates artificial astrocytes into MLP. We showed how astrocytic noise works better than gaussian noise and how synaptic plasticity could be driven by astrocytes. We proposed technique for optimizing astrocyte weight that yields better results than the original model. Since all presented techniques have been shown to help avoiding being stuck in local minima, it is definitely possible to attempt applying them to different models that have this tendency. Therefore future research could focus on integrating these techniques into recurrent, self-organizing or spiking models.

5 Acknowledgment

This work was supported by the grant APVV-0668-12 and KEGA grant 017UK-4/2016.

References

- Alvarellos-González, A., Pazos, A. and Porto-Pazos, A. B. (2012). Computational models of neuron-astrocyte interactions lead to improved efficacy in the performance of neural networks. *Computational and Mathematical Methods in Medicine*, 2012.
- Brightman, M. and Reese, T. (1969). Junctions between intimately apposed cell membranes in the vertebrate brain. *Journal of Cell Biology*, 40(3):648–677.
- Hilgetag, C. C. and Barbas, H. (2009). Are there ten times more glia than neurons in the brain? *Brain Structure and Function*, 213(4-5):365–366.
- Ikuta, C., Uwate, Y. and Nishio, Y. (2010). Chaos glial network connected to multi-layer perceptron for solving two-spiral problem. In *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1360–1363. IEEE.
- Ikuta, C., Uwate, Y. and Nishio, Y. (2011). Performance and features of multi-layer perceptron with impulse glial network. In *Proceedings of the International Joint Conference on Neural Networks*, pages 2536–2541. IEEE.
- Irizarry-Valle, Y. and Parker, A. C. (2015). An astrocyte neuromorphic circuit that influences neuronal phase synchrony. *IEEE Transactions on Biomedical Circuits and Systems*, 9(2):175–187.
- Irizarry-Valle, Y., Parker, A. C. and Joshi, J. (2013). A cmos neuromorphic approach to emulate neuro-astrocyte interactions. In *International Joint Conference on Neural Networks*, pages 1–7.
- Joshi, J., Parker, A. C. and Tseng, K.-C. (2011). An in-silico glial microdomain to invoke excitability in cortical neural networks. In *IEEE International Symposium of Circuits and Systems*, pages 681–684.
- Mesejo, P., Ibáñez, O., Fernández-Blanco, E., Cedrón, F., Pazos, A. and Porto-Pazos, A. B. (2015). Artificial neuron–glia networks learning approach based on cooperative coevolution. *International Journal of Neural Systems*, 25(04):1550012.
- Murray, A. F. and Edwards, P. J. (1994). Enhanced mlp performance and fault tolerance resulting from synaptic weight noise during training. *IEEE Transactions on Neural Networks*, 5(5):792–802.
- Orkand, R., Nicholls, J. and Kuffler, S. (1966). Effect of nerve impulses on the membrane potential of glial cells in the central nervous system of amphibia. *Journal of Neurophysiology*, 29(4):788–806.
- Oschmann, F., Berry, H., Obermayer, K. and Lenk, K. (2017). From in silico astrocyte cell models to neuron-astrocyte network models: A review. *Brain Research Bulletin*.
- Volman, V., Bazhenov, M. and Sejnowski, T. (2012). Computational models of neuron-astrocyte interaction in epilepsy. *Frontiers in Computational Neuroscience*, 6:58.
- Wade, J., Kelso, S., Crunelli, V., McDaid, L. J. and Harkin, J. (2014). *Biophysically based Computational Models of Astrocyte ~ Neuron Coupling and their Functional Significance*. Frontiers E-books.

Vnímání pohybu: Reflektovaný a nereflektovaný pohled v souvislosti s architekturou

Ivan M. Havel a Monika Mitášová

Centrum pro teoretická studia při Univerzitě Karlově a AV ČR, Jilská 1, 110 00 Praha 2, ČR
Katedra teorie a dějin umění, Trnavská univerzita, Hornopotočná 23, 918 43 Trnava, SR
havel@cts.cuni.cz, mitasovamonika@gmail.com

Abstrakt

I když se domy a města v obvyklém smyslu nepohybují, mohou pohyb určitým způsobem vyjadřovat, popřípadě vytváret prostor k lidskému pohybu a jednání. Příspěvek se zabývá otázkou, jak lze z fenomenologického hlediska vnímat místní pohyb lidí a věcí v prostoru, reflektovat pohyb vlastní a jak lze vlastním pohybem prostor vnímat. S odkazem na současné architektonické projekty se zamýšíme, zda je možno pohyb a jeho vnímání zaznamenávat, projektovat, ba i stavebně realizovat.

1 Co se hýbe?

Kromě *prostoru* další dva nejužívanější a přitom nejméně vyjasňované termíny ve vědě, filosofii, ba i v běžném životě, jsou pojmy *čas* a *pohyb*. Čas si lze jen těžko představit bez pohybu a pohyb bez času, velmi však záleží na kontextu, v jakém je to či ono slovo užíváno. V této práci se zaměříme na pohyb, a to spíše z hlediska jeho vnímání, čili na otázku lidské percepce pohybu. Speciálně nám přitom půjde o subjektivní vnímání, čili (v jazyku kognitivní vědy) o perspektivu „*první osoby*“.¹ Jinými slovy bude nás zajímat, jak pohyb může být vnímán či prožíván a popřípadě také konceptualizován člověkem. Specificky pak člověkem, který se ocítá v urbánním a architektonickém prostředí, v prostoru architektury, přičemž cosi z našich úvah se snad může stát inspirací také v jiných vědeckých a uměleckých kontextech.

Vyjděme z celkem neproblematického předpokladu, že člověk má zpravidla něco, co nazveme *smysl pro pohyb*. Tím ovšem nemáme na mysli nějaký speciální smyslový orgán (jakoby „*šestý smysl*“), který svého nositele zpravuje o aktuálních vnějších nebo jeho vlastních pohybech.² Spíše nám

jde o obecně chápou lidskou zkušenosní dispozici, způsobilost či schopnost rozpoznávat pohyb a rozumět mu (sousloví „*smysl pro...*“ zde má tedy podobný význam, jako ve frázi „*smysl pro humor*“ – ten mohu mít i když zrovna není čemu se smát. Později se zmíníme ještě o dalším „*smyslu pro...*“, totíž pro estetiku).³

Rozpoznávání pohybu člověkem předpokládá, že existuje perceptivní i konceptuální rozdíl mezi tím, *co* se pohybuje (nebo o čem uvažuji, že se to pohybuje) a tím, *vůči čemu* se to pohybuje. Tak třeba letící pták vůči obloze; šmátrající ruka vůči zdi; ten/ta, kdo stoupá po schodech, vůči schodišti pod nohami. V takových případech lze mluvit o *referenčním poli* pohybu, ve kterém lze pohyb vnímat jako vztah *popředí* (figury) a *pozadí* (oblohy, zdi, schodiště...). Pohyb sám ovšem nemusí být vnímán jen vizuálně, ale též hmatem, sluchem, kinesteticky apod., nicméně se ted zaměříme hlavně na vizuální vnímání, lépe řečeno na *pohled* (k haptickému vnímání pohybu, kinestezi a ke vnímání pohybem se vrátíme později).

Při aktuálním prožitku pohybu nemusí být rozlišování mezi tím, kdy je vnímán v popředí a kdy v pozadí jednoznačné. V jedoucím vlaku se mohu soustředit na to, co je před oknem, zatímco scéna za oknem rozmazeně ubíhá. Jindy se naopak soustředím na scénu za oknem a dobré vím, že co se pohybuje, jsem já sám a můj vlak. Snad nemusíme připomínat často vzpomínaný pohyb Slunce po obloze.

Pokud porovnáme to, jak se pohyb může jevit subjektu v perspektivě první osoby, s tím, co se ví o aktuální skutečnosti (čili v perspektivě „*třetí osoby*“), můžeme v prvním přiblížení rozlišovat pohyb *aktuálně vnímaný* od pohybu *virtuálního* (tj. od pohybu jen v představě, např. při pohledu na schody anebo na střechy domů). Lze také rozlišovat pohyb *zdánlivý, iluzivní* (jen se mi něco jeví jakoby v pohybu)⁴ od pohybu *vyvozeného (dedukovaného)*

¹ Chceme-li v následujícím textu zdůraznit tuto perspektivu, uchylujeme se k formulacím v první osobě jednotného čísla. S počtem autorů to nesouvisí.

² Rozumí se pohyb *místní* (mění se poloha něčeho), nikoliv pohyb v obecném smyslu (jakákoliv změna). Z kontextu bude jasné, zda přitom jde o pohyb aktuální,

anebo jen myšlený či tušený. O vnímání pohybu z hlediska psychologického viz Sikl (2012), kap. 6.

³ O kategorii „*smyslů pro (něco)*“ viz též kapitolu „*Smysly a smysly*“ v knize Havel (2013), s. 233–240.

⁴ Speciálním případem by byl pohyb vnímaný ve snu – ten lze považovat buď za reálný (tak je pociťován během

na základě zkušenosti a znalosti světa (pohyb motýla na nástenné malbě v místnosti). Z jiného hlediska lze rozlišovat pohyb *věcný*, pohyb *můj vlastní* a pohyb *jiných* osob.

V míře, v jaké se zajímáme o lidské vnímání pohybu, nelze si nevšimnout mlhavosti hranic mezi aktuálně vnímaným pohybem (mějme na mysli vizuální vnímání a věcný místní pohyb) a pohybem z hlediska člověka nevnímatelným – na jedné straně to může být pohyb tak rychlý, že jej neumíme zaregistrovat (letící projektil), na opačné straně pohyb tak pomalý, že se nám jeví jako nehyb (malá ručička hodinek, pohyb geologických vrstev). Tyto hranice lze pochopitelně zkoumat experimentálně (případně introspekčně), přičemž jejich existence a mlhavost zůstává nedotčena.

Příbuzným širším tématem jsou *mžikové zážitky*. Na velmi krátký okamžik se rozsvítí světlo (nebo otevřou oči) a otázkou je, zda si v prostředí, které se přítom objeví, stihнемe vůbec všimnout nějakého pohybu. I toto lze experimentálně (či introspekčně) zkoumat a lze nabídnout celkem snadnou hypotézu, že rozsah vnimatelnosti je tím užší, čím kratší je trvání mžiku a čím méně výrazný je rozdíl mezi tím, co se pohybuje, a jeho pozadím.

2 Reflexe pohledu na pohyb

Začneme příkladem. Stoupám po schodišti do vyššího patra domu a mým hlavním zájmem je (třeba) to, co mě v onom vyšším patře čeká. Pokud schodiště dobře znám, anebo je navrženo a postaveno stejně jako převážná většina mně známých schodišť, neuvědomuju si příliš svůj vlastní pohyb, ba ani skutečnost, že se vůbec nalézám na nějakém schodišti. To vše je skryto v pozadí mé uvědoměné pozornosti, i když to je v mém zorném poli. Poněkud neobratně řečeno, svůj pohyb cítím a vnímám, dokonce okrajově i *vidím*, avšak *nereflektuji*.

Když nekráčím přímo za tím, co mě čeká ve vyšším patře, anebo se ocitnu na neznámém schodišti (v hale, v jednotlivých místnostech a postupně v celé budově) více zaměruji pozornost na svůj pohyb a na můj způsob tohoto pohybu. Například se soustředím na to, jak stoupám po stupních visutého, zavěšeného schodiště a očima si vybíram vhodné místo, kam stoupnout právě zde na těchto podivuhodných, pro mne nezvyklých schodech (anebo nezvyklých obecně v naší kultuře). Právě vnímaná diference procházky schodištěm, anebo tohoto, řekněme atypického schodiště, od jejich obvyklého průběhu anebo podoby (sklon, rozměry stupňů, tvar madla) koriguje, anebo ruší konvenci, ustavuje odchylku popřípadě stabilizuje konvenci novou. Ale tím také testuje a obnovuje smysl pro pohyb: obohacuje jej – přináší nový

snu), anebo za zdánlivý (což si uvědomíme až po probuzení).

zážitek, novou zkušenosť pohybu. Otázkou je, jakou úlohu při tom hraje reflexe: reflexe vlastního pohybu a reflexe schodiště (haly, domu) v tomto vlastním pohybu.

Smysl pro pohyb a obecnou zkušenosť pohybu lze chápát i jako jakési *předporozumění* pohybu aktuálně prožívámu. Svůj smysl pro pohyb a předporozumění mu člověk průběžně testuje a koriguje ve vztahu ke způsobům vnímání aktuálního pohybu. Testuje je konkrétním zřetězením, sledem vjemů ve *spojitém* pohybu (krok za krokem), ale také v pohybu *přerušovaném* zastaveními. Ze smyslu pro pohyb obecně a z předchůdné (optické, sluchové, haptické, motorické) zkušenosti se prostřednictvím spojitych i přerušovaných pohybů a zastavení (jako jsou pohledy zdálky a zblízka, posuny po schodišti v prostoru anebo *podél zdi* a doteky madel v *prostoru*, *na zdi* anebo v *nice zdi*) mění orientace a reorientace v prostoru a čase pohybu. Nové časoprostorové rozvrhování vjemů schodiště, domu, ulice, města (haptickotaktilní prostor ve vztahu k blízkému i vzdálenému vizuálnímu prostoru) a vše se koriguje koeficientem posunu, změny: probíhá kontinuální variování orientací a reorientací. Pokud bereme v úvahu referenční pole pohybu v architektuře, lze si klást otázku, kdy a jak člověk v pohybu vnímá schodiště, dům (architekturu) především jako vztazný rámec anebo vztaznou soustavu (směry, rozměry) svého pohybu, kdy se dostává spíše do pozadí a kdy vystupuje do popředí. Ale úvodní otázkou by mohlo být, co nám dovoluje architekturu vůbec vnímat jakožto architekturu.

Historik umění Alois Rieg (1893, 1901) uvažoval o *estetické dispozici* lidí. Obdobně, jako jsme mluvili o smyslu pro pohyb, považoval tuto dispozici za jisté duchovní naladění, jež má nevědomou, instinktivní povahu. Právě tato dispozice zakládá vztah člověka k jeho okolí a ke světu. Díky ní je možný estetický vjem přírodních a uměleckých jevů, který se rodí mimo jiné v polaritách haptického a optického, blízkého a vzdáleného pohledu. Estetickou dispozici sdílí publikum i tvůrci umění, ale umělecké jevy vznikají podle Rieglia ve vztahu k nadindividuální *kunstwollen* – k vůli k umění. Rieglova koncepce umění (a uměleckých stylů) směřuje od haptického podání uměleckých předmětů k jejich optickému podání, ostatně západní umění obecně směřuje od hmatovosti k zrakovosti (v jednotlivých kulturách a historických obdobích mohou přitom tyto polarity nabývat různou váhu, rozmanité vzájemné vztahy, smysl, významy a hodnoty). Hmatové podání u Rieglia znázorňuje věci „bez prostoru a času“ zatímco u optického podání je tomu naopak: přítomnost času a prostoru není ze znázornění věci vyloučena. Prostor a čas jsou samy nevěcné, ale umožňují přístup k věcem: poukazují na jejich vzájemnou blízkost anebo distanci. Pro povahu umění (a jeho stylů) je tedy důležité, zda se čas

a prostor v haptickém názoru či znázornění skrývají, anebo se naopak v optickém podání objevují. V tomto smyslu mohou jednotlivé styly u Rieglu ukazovat na rozmanitá pojednání časoprostorové povahy bytí člověka ve světě. A v tomto smyslu zakládají také rozvrh, strukturu popřípadě formu výrazu uměleckého díla.

V uměleckém díle se dotykový styl tvorby umění projevuje jeho plastičnosti – Riegl ho považuje za *plasticky objektivní*, je založen na *objektivním pohledu zblízka*. Zrakový styl tvorby umění je založen na *opticky subjektivním pohledu zdálky*. S těmito kulturními a dobovými preferencemi umění pak také souvisí například způsoby situování obyvatel (subjektů) v prostoru architektury – a souvisí s tím také její „látková uzavřenosť“ anebo „optické propojení“ s okolním prostorem... Z toho vyplývají i haptické anebo optické preferenze v utváření a recepci architektonických prostorů u těch kulturních komunit, které jsou v určitých dobách nositeli nadindividuální *kunstwollen* – vůle zakládající dynamické napětí, výměny mezi plastickými výrazovými formami a stylovými principy (antický Řím) na jedné straně a stylovými principy a výrazovými formami optickými (Germáni) na straně druhé. Riegl vypracoval „genetické řady“ stylových forem, u nichž chronologicky pozdější stylová forma může být geneticky mladší a naopak. Nabízí tedy vlastní, *imanentní vývoj uměleckých výrazových forem* nezávislý na mimouměleckých jevech. V tomto jeho modelu se uměním může stát jenom to, co zasahuje formu uměleckého díla a projevuje se v nich, tedy pokud slouží jejich výrazu.

Ale Rieglova polarita optického a haptického, blízkého a vzdáleného pohledu v umění nadále inspiruje i mimo systém, který vytvořil. Z této polarity vycházeli a přehodnotili ji také Deleuze a Guattari (2010, s. 563–572), když uvažovali o povaze a vztazích dvou prostorů v umění: jednak prostoru *hladkého* (spojitého, blízkého a „absolutně lokálního“) a jednak prostoru *rýhovaného* (nespojitěho, ubíhajícího do dálky a „relativně globálního“). Mezi nimi zavedli prostor třetí: *prostor procesu, přechodu (stávání se)*, který s oběma zmíněnými rieglovskými polaritami nakládá velmi volně. Jako příklady hladkého prostoru v přírodě uváděli: moře, poušť, led, v kultuře například orientální umění. Znázornění hladkého prostoru nepoužívá pohled z dálky, obraz je tvořen zblízka. Nerozlišuje „vnitřek“ ani „vnějšek“ jelikož nepracuje s mezerami, intervaly, ani otvary, není v něm rozlišené popředí ani pozadí, orientace v něm je založena především na taktilních vztazích: také oko se v takovém blízkém, lokálním prostoru „dotýká“ toho nač hledí a „duch“ je v takovém prostoru vlastně „prstem“. V tomto smyslu haptický režim zobrazování neosahuje ani pozadí, ani roviny, ani obrys. Obsahuje změny směru (orientace) a spoje lokálních částí.

Naproti tomu znázornění rozčleněného (mnohosměrně rýhovaného) prostoru je určováno pohle-

dem z dálky, má optickou perspektivu: rozlišují se v něm distance, roviny, objemy a orientační body. Ale obsahuje také křížení rovin. Jako příklad v kultuře Deleuze s Guattarim uváděli severské umění. Linie v něm má převažující orientaci, je „přímočará“ v tom smyslu, že vede od bodu k bodu (spojuje dva body, figury, obrysy) a vymezuje obrys, zatímco linie v hladkém, haptickém prostoru má „mnohonásobnou orientaci“, vede mezi body (figurami, obrysy) a nevymezuje, nevytyčuje žádný obrys – odchyluje se, mění směr – je to kontinuální variace. Nemá rieglovskou stabilizovanou *formu výrazu*, ale má *sílu výrazu*. A tak jako síla má také opakování (s variací). Tomuto opakování (s variací) odpovídá diagram, zatímco stabilizované formě výrazu by odpovíděl nákres a realizační výkres. Linie jako diagram síly kontinuální variace *navozuje přechody* mezi haptickým a optickým, pohledem z blízka – pohledem z dálky, pohledem na zeď – pohledem na prostor, protože její působení v prostředí je organické. Analogické vlastnosti mají i kulturní prostředníci mezi orientem a severem: nomádské umění a jeho následovníci: barbarské, gotické a moderní umění znázorňující a uchopující přechody, procesy a události. Otázkou je, co zde navozuje a může navozovat reflexi pohledu-doteku anebo „dotýkajících se pohledů“ (haptické optičnosti) a „vidících doteků“ (optické hapticity).

Důležitá je také otázka, jaká situace nastane, když aktuální haptické, optické (hapticko-optické) i motorické vjemy pohybu rozdílují automatismy mého pohledu a dovolují mi chápout smysl schodiště rozmanitými způsoby, také alternativně anebo nově (například při pohledu z dálky jako prostor rýhovaný jednotlivými stupni nahoru/dolů, při pohledu z blízka jako zřetězení, sérii přechodů mezi lokálními horizonty mého prostoru (v) pohybu, anebo jako prostor pro zastavení, hru a výhledku do dalších prostor domu...). Například vnímání schodiště jako řetězce hladkých přechodů od předchozího „lokálního horizontu“ k následujícímu mi umožňuje novou orientaci a aktuální rozvrhování toho, co je „dole“, „nahore“, „vpravo“ a „vlevo“, a to díky pohybu a krátkým zastavením v něm, kdy nevnímám žádné stabilní, usazené „popředí“ ani „pozadí“ referenčního pole.

V tomto smyslu nastává cosi jako destabilizace, otřes mých předchůdných rozvrhů schodiště, kdy ho na základě pohledu z dálky vidím jako jedno těleso v popředí anebo v pozadí vstupních prostor domu, jako jeden rýhovaný prostor s intervaly mezi stupni. Haptické vjem (madla, zdi) zase umožňují, že sled z blízka viděných „lokálních horizontů“ prožívám a rozumím mu jako jednomu spojitému, hladkému celku s hloubkou... Motorické vjem dovolují chápout přechod a průběh (posun, klesání, stoupání) a následnost jednotlivých zastavení (schodů a podešti – vyhlídkových ploch schodiště určených k zastavení a obratu, ostatně původní význam slova reflexe je: *obrat*). Otázka je,

jak se tyto „obraty“ vnímání od pohybu k objektu (schodišti a jeho architektuře) a zpět podílejí a mohou podílet na oscilaci mezi užíváním schodiště a jeho estetickou či architektonickou reflexí, respektive interpretací. Na jedné straně již dynamika vjemů – způsob jakým moje vjemy rozbíjejí obraz-objekt schodiště a jak jej na druhé straně nově ustavují, rozvrhují a uspořádají – předpokládá spolupodílení mojí estetické dispozice. Na druhé straně rytmičnost a nerytmičnost chůze po schodech, její neusporádanost a uspořádání, pohyb a zastavení – ale také obraty na zlomech schodiště – působí jako události a tedy jakési síly, které testují a korigují proto-estetické a protoarchitektonické prožitky a koncepty. Vyvolávají mizení etablovaných estetických (a architektonických) konvencí při vnímání obrazů resp. objektů a vedou k množství dosud neetablovaných konvencí. Tyto síly mi umožňují rozmanitě chápat a různě interpretovat také rozvrh a strukturu schodiště, po kterém kráčím a které bylo navrženo a vyrobeno určitým kulturním okruhem v jisté době jako výraz jeho řemeslného a stavebního úsilí, anebo jako výraz jeho *kunstwollen*, respektive jeho koncepce prostoru a architektury.

3 Pohlížení na pohled

Na začátku minulé kapitoly jsme nabídli příklad domu se schodištěm, po němž se pohybují fyzicky. Nyní předpokládejme, že se fyzicky nepohybují, nýbrž že stojí pod schodištěm a pouze na ně hledíme. Zjišťují, že je to opravdu schodiště – takové přece znám a vím, k čemu je dobré. Mnohokrát jsem chodil nahoru dolů po podobných schodištích, můj pohled je banální (automatický), není tu důvod k nějaké jeho reflexi. Ale zamysleme se: jak vůbec vím, že co vidím, je schodiště, něco, po čem se dá stoupat nebo sestupovat? Odpověď zní: *mé tělo to ví*. Ví to opticky, hapticky, motoricky. Když v mládí se to naučilo, postupně si na podobná schodiště zvyklo, tak-říkajíc „vtělilo“ (anebo „privtělilo“) si je. Já už o tom vlastně moc neví, teď prostě jen *vidím*, že to je schodiště. K tomu stačí mít je v zorném poli, byť třeba jen okrajově. Není nutné je pečlivě zkoumat a stejně tak není nutné reflektovat akt vlastního pohledu.

Je příznačné, že v tomto smyslu lze mimo-esteticky, ale také proto-esteticky, spatřit schodiště *jakožto* schodiště i při velmi krátkém, mžikovém pohledu. Při něm vlastně ani nerozliším, zda hledím na reálné schodiště nebo třeba na sgrafito se schodištěm namalovaným na zdi domu. Při mžikovém pohledu není dost času na vybíráni vhodného místa, kam stoupnout. Ale přesto je čas a prostor na vznik vjemu, jenž může být reflektován v kontextu estetické dispozice a interpretován v kontextu architektonických koncepcí prostoru.

V případě věcného pohybu nebo pohybu jiné osoby lze otázku vnimatelnosti a reflexe obrátit: lze

„vidět“ pohyb, který je pouze zobrazen? Co však znamená „zobrazený pohyb“? Jedna možnost je *fázování*: vidíme a zobrazujeme sled zastavených obrazů v jakém-si intervalu vzájemných posunů, změn místa popřípadě místa a času. Existují rozfázované fotografie Etienna Julese Mareyho z 80. let 19. stol., a známý je v tomto směru také například obraz Marcela Duchampa *Akt sestupující se schodů č. 2* (1912), který nám svým fázováním pohybující se postavy dává najevo nejen možnost sledovat pohyb (a obraz pohybu), ale také možnost položit si otázku, jak vůbec lze pohyb vidět a zobrazenit v malířství 20. století. Duchampův provokativní a paradoxní výrok napsaný na lístku zařazeném do objektu *Zelená krabice* z roku 1934 o tom, že „*je možno pohlížet na pohled, ale není možno slyšet slyšení*“⁵ ukazuje, že uvažoval nejen o opakování, ale také o odložené, odsunuté (malířské i mimomalířské) reflexi pohledu na záznamy pohybu i nehybu. Lze to doložit jeho „galeriem“, což jsou krabice či kufříky plné fotografií, grafik a textů, které bylo možno opakován rozložit, prohlížet, zapsat své dojmy a zase složit a vrátit zpět.

4 Prostor k pohybu

Hledím na schodiště a nevidím žádný pohyb... Nikoliv! Já přece „vidím“ nejen schodiště, ale jaksi předběžně i jakousi pomyslnou dráhu, diagram (ze vzpomínky či předjímání) svého *vlastního pohybu* na něm, jak po něm stoupám či sestupuji. Označili jsme to jako pohyb *virtuální*. Schodiště mi totiž otevírá *prostor k pohybu v místě i čase* (srov. Havel a Mitášová, 2004).

Je vhodná chvíle se na věc podívat trochu obecněji. Představme si člověka, který se nachází v prostoru nějaké stavby se stěnami, chodbami, schodišti, terasami a s průhledy a průchody do dalších místností. Zajímá nás, zda a jak dotyčný člověk prožívá tento prostor ze svého hlediska – čili v perspektivě první osoby – *jakožto prostor k pohybu*. Jde o aktuální či potenciální pohyb tělesný, a to *můj vlastní pohyb* v tom smyslu, že je v *mé moci*.⁶

V jistém smyslu jako by mě strukturovaný prostor k pohybu přímo vybízel. Pravda, je tu sice možnost jej interpretovat různě, třeba geometricky (a tudíž objektizovaně) nebo esteticky, avšak nám zde jde právě o ono *vybízení* k pohybu. K tomu patří i to, že jaksi předvedomě vím, že mám více možností na výběr: mohu se třeba odebrat jedním z průchodů do další místnosti, mohu vystoupit schodištěm do vyššího podlaží a pokračovat na terasu. Má i jedno zásadnější rozhodnutí: bud' se chystat pohyb uskutečnit, anebo mi stačí pouhý pohled, v němž jsou ony prve zmíněné volby implicitně obsaženy.

⁵ „One can look at (see) seeing, / one can't hear hearing.“ Viz Sanouillet, M. a Peterson, E. (Eds.). (1973).

⁶ Viz též pozn. 1 výše.

Jak vidíme, lze na prostor k pohybu v nějakém prostředí buď pohlížet fenomenologicky, jako na rozvinuté pole možností subjektu který se pro pohyb či jiné jednání rozhoduje, anebo můžeme totéž prostředí popisovat objektivizovaným jazykem geometrie, estetiky atd. První pohled je ve vědě často zanedbáván, i když lidské zaujetí a prožívání je existenciálně prvotní než geometrický náhled. Na straně subjektu se zaměříme jen na ten druh lidského zaujetí, které bezprostředně souvisí se senzomotorickým poznáváním těch stránek okolního prostředí, které pohyb umožňují a omezují. (Proto v této práci odhlížíme od rozmanitých individuálních stavů a záměrů subjektu – snad kromě estetické dimenze a smyslu pro pohyb –, jakkoliv by mohly být psychologicky zajímavé i jiné nálady a naladění v konkrétních případech).

Právě pohyb je to, co nám umožňuje pojímat prostor nejen skrze jakoukoliv vlastní tělesnost, ale skrze kinestesi: „*Jenom proplétání těla prostorem (...) může spojit prostor, tělo, oko a ducha,*“ píše Steven Holl (2003, s. 38).

5 Pohybem k prostoru aneb Vnímání pohybem

V textu *Architektura pohybu* (*Architecture of Movement*) se Patrik Schumacher z ateliéru nedávno zesnulé architektky Zahy Hadid táže, zda může existovat taková teorie anebo způsob koncipování prostoru, které přesahují „arch-architektonický“⁷ (první anebo prvotní architektonický) prostor který by se neřídil modulací (nebyl utvářen s ohledem na apriorní modulový rád). Má na mysli (v deleuzovské terminologii procesuální, přechodový) prostor, který by byl definován jen pohybem samým⁸ bez nezávislého anebo apriorního referenčního systému. Jinak řečeno, ptá se, jak lze například navrhnut systém pohybu po budově (příp. ve městě) bez toho, aby byly předem známy body, které má tento systém komunikací propojovat. Takový prostor by se podle Schumachera spíš podobal prostoru „tance“ než prostoru nejkratších spojnic mezi předem známými stanovišti. Tedy pokud si pod „prostorem tance“ nepředstavujeme konkrétní choreografii ani taneční figury, ale prostor, který poskytuje možnost tvůrčího pohybu obecně, na rozdíl od prostoru pohybu utilitárního anebo čistě funkčního. Jak by takovýto „tanec“ mezi moduly mohl například vymezovat prostor bez karteziánského geometrického rastru?

V této souvislosti nelze nezmínit (a Schumacher tak činí) Baudelaireova tuláka (*flaneur*)

⁷ „Can there be a theory or a conception of space beyond the "architectural" space of modular control?“ Schumacher (1996), s. 134-135.

⁸ „[D]efining space through movement alone, without an independent and prior system of reference?“, Ibidem.

jehož pohyb vede spíše touha než cokoliv jiného a rekonfiguruje prostor města 19. století anebo takzvané „konstruované situace“, jejichž autory byli členové Situacionistické internacionály (založené v roce 1957), jejichž cílem bylo zmnožování „poeticích subjektů a objektů“ a pořádání *her* poeticích subjektů s poetickými objekty: „*Naše situace budou efemérní, bez budoucnosti. Průchody?*“⁹ Tyto prostory hry ve městě měly rušit, překonat umění jako specializovanou činnost určenou pro galerie a trh, a to tím, že se uskutečňují v čase a prostoru každodenního života, a navíc, že tento každodenní život „přetvářejí“ v duchu poeticích zákonitostí. Umělecká transformace každodennosti nastává v „konstruované situaci“ a díky ní – je to ten „*moment života, který je konkrétně a záměrně konstruován kolektivní organizací jednotného prostředí a hry události*“.¹⁰

K vytváření těchto situací umělci a architekti Situacionistické internacionály převzali a rozvíjeli postupy Lettristické internacionály (v letech 1952-57) nazvané *dérive* (unášet – ve smyslu nechat se unášet) a *détournement* (vychýlení něčeho). *Dérive* byl způsob experimentálního chování, byla to „*technika rychlého průchodu skrze rozmanitá prostředí*“. Počas *dérive* „*se jedna či více osob na určitou dobu vzdávají svých vztahů, své práce a aktivit ve volném čase, jakož i všech dalších obvyklých důvodů k pohybu a činnosti, a nechávají se vést přitažlivostí terénu a toho, s čím se v něm setkávají.*“¹¹ *Dérive* byla analogie

⁹ „We have to multiply poetic subjects and objects — which are now unfortunately so rare that the slightest ones take on an exaggerated emotional importance — and we have to organize games for these poetic subjects to play with these poetic objects. This is our entire program, which is essentially transitory. Our situations will be ephemeral, without a future. Passageways.“ Debord, G.: Report on the Construction of Situations and on the International Situationist Tendency's Conditions of Organization and Action. 1957 (anglický překlad Ken Knabb), dostupné na webu www.cddc.vt.edu/sionline.

¹⁰ „A moment of life concretely and deliberately constructed by the collective organization of a unitary ambiance and a game of events.“ In: „Definitions“, Internationale Situationnist, č. 1, 1958 (anglický překlad Ken Knabb), dostupné na webu (ibidem).

¹¹ „A mode of experimental behavior linked to the conditions of urban society: a technique of rapid passage through varied ambiances. The term also designates a specific uninterrupted period of deriving.“ In: Ibidem. „In a *dérive* one or more persons during a certain period drop their relations, their work and leisure activities, and all their other usual motives for movement and action, and let themselves be drawn by the attractions of the terrain and the encounters they find there. Chance is a less important factor in this activity than one might think: from a *dérive* point of view cities have psychogeographical contours, with constant

procházky bez předem vypracovaného plánu a cíle: znamenala toulání se, „bloumání“, při němž bylo možno objevovat i známá městská prostředí v nových souvislostech a nově. Z reflektovaných vjemů, emocí, a také zaznamenaných prožitků byly sestavovány „*psycho-geografické*“ mapy města. Působení *dérive* bylo někdy různými způsoby zesilováno: třeba falšováním jízdních rádů nebo označováním některých ulic paradoxními výzvami. Výzva: „*Už nikdy nepracujte*“ osloivila i revoltující studenty pařížského jara 1968. *Détournement* znamenal „*vychýlení již existujících estetických prvků*“ a „*včlenění současné nebo starší umělecké produkce do celkové konstrukce prostředí.*“¹² Cílem tedy nebyl vznik situacionistické malby, koláže, ani urbanismu či architektury, ale jejich „situacionistická“ reflexe, interpretace a reinterpretace mimo jejich každodenní užití.

Schumacher tyto strategie považuje za předchůdce současné *architektury pohybu*, kteří rozvrátili představy o hierarchizovaném a teritoriálním prostoru architektury. Také díky nim Schumacher určil a charakterizoval současný architektonický prostor generovaný „spontáním pohybem“. Nazval jej „anti-Architekturou“ ve smyslu anti-modulárního (anti-rýhovaného) prostoru v prostoru modulárním a tedy rýhovaném a uvedl dvě jeho krajní polarity. Nizozemský ateliér UNStudio navrhl první polaritu jako prostor s geometrií doupěte či nory (*burrow-geometry*), tedy prostor jako díry v modulové architektonické struktuře. Ta obsahuje mimo jiné i prostory labyrintické a meziprostory, obdobné, jaké navrhl ve svých modelech holandský výtvarník Constant jako situacionistickou interpretaci propojeného „celosvětového města“ – *Nového Babylonu* (1959–1974). UStudio navrhlo děravý prostor v modulární struktuře – průchozí prostory – tak, že se objevují jako „překvapení“ ve výškových bytových stavbách, které jsou modulární („rýhované“) a lze je chápat jako zvláštní extenze architektury vnímané v pohybu, například na procházce.

Druhou krajní polaritou současné architektury pohybu je vila – spirála v modulové kostce od ateliéru Zahy Hadid (spirála u Deleuze a Guattariho představuje jeden z typů „hladkého“ prostoru). Aby takové prostory vznikly, bylo nutno virtuálně generovat pohyb nějaké částice/tělesa v modulárním systému. Analogické projekty navrhoval v 90. letech i americký architekt Greg Lynn. Ale jeho laboratorium robotické architektury na UCLA v Los Angeles v současnosti testuje a navrhuje nejen stavby, které jsou v reálném čase rekonfigurovány

currents, fixed points and vortexes that strongly discourage entry into or exit from certain zones.“ In: Debord, G.: Theory of the Dérive, *Les Lèvres Nues*, č. 9, reprinted in *Internationale Situationiste* č. 2, (angl. překlad Ken Knabb), dostupné na webu (ibidem).
¹² Dostupné na webu (ibidem).

a přeupořádávány roboticky (a tedy celé jejich moduly a segmenty se přesouvají a přeskladají), anebo domy – robotické mobily, ale testuje také navrhování pomocí vlastních gest a pohybů v prostoru a čase rozšířené (*augmented*) reality. Tím se otevírá zcela nové téma: nejen vnímání a reflexe architektury v pohybu, nejen počítavě-robotické navrhování současné architektury pohybu, ale také současná architektura navrhovaná a reflektovaná pohybujícím se architektem: gesticky, choreograficky. A otevírá se tedy také otázka: co je *chora* tohoto nového choreografického navrhování – organizace prostoru – pohybem těla?

6 Závěrem trochu spekulace

Zatím jsme se zabývali tématem pohybu a pohledu na něj, ať už teoreticky, či s odkazem na reálné příklady. Přitom jsme se soustředili na pohyb člověka nebo věci v prostoru takříkajíc mesoskopickém – typickém pro pohybování člověka z místa na místo. Hlavní inspirací proto byla pro nás architektura, a to v tom smyslu, že nejde jen o budovy samy o sobě, ale i o jejich projektování, užívání, reflexi v čase a užívání jako prostředí k lidskému pohybu. Architekt navrhuje nějaké uspořádání (strukturu, prostor) a přitom si musí (případně za použití počítavé animace) ve své mysli představovat, jak jeho dílo bude postaveno, vnímáno a užíváno hypotetickým obyvatelem či návštěvníkem.

V nedávné době byly navrženy nové prostředky k tomu, aby architekt mohl oscilovat mezi geometrií a reálným prostorem a stejně tak mezi idejí a vizualizací navrhovaného prostoru. Nepřekvapuje, že se přitom počítá s nejnovějšími možnostmi digitální technologie, jako je vizualizace ve 3D, interaktivita, virtuální a rozšířená realita, počítavová navigace a jiné. Oba autoři jsme teoreticky zaměření, a proto nabízíme závěrem dosti spekulativní uvažování. Nebylo by však překvapivé, kdyby se leccos z toho, co je nyní ve sféře idejí, představ a fantazií (sci-fi nevyjímaje), stalo realitou už v době, kdy budete tyto úvahy číst.

Jeden příklad. Architekt má za úkol navrhnut rekonstrukci objektu (například opuštěné továrny v Detroitu, jakou Greg Lynn¹³ vystavoval na Bienále architektury v Benátkách 2016) a přitom zachovat některé jeho prostorové nebo tvarové prvky a navrhnut nové. Obvykle by se všechno muselo vyměřovat a zanášet do výkresů, které by pak architekt překresloval (byť nejspíš s „pasivní“ pomocí počítače), aniž by se mohl opřít o jakýkoliv jiný, než nakreslený anebo počítačem zobrazený vzhled co do toho, jak se jeho nápady budou do původního prostoru včleňovat. Uvažujme však, že si místo toho nasadí speciální holografické brýle,

¹³ Viz <http://worldarchitecture.org/articles-links/cemch/greg_lynn_transforms_detroit_car_factory_microsoft_hololens_at_venice_biennale.html>.

které budou kombinovat aktuální a virtuální realitu a navíc budou interaktivní. S nimi se vydá do původního objektu (továrny), kde si zkusmo bude vymýšlet nové alternativní prvky nebo prostorové formy. Ostatně ty třeba může mít připravené v počítači, aby mohl mezi nimi volit. Tyto prvky nebo formy se mu budou ve virtuální podobě ihned promítat přes aktuálně přítomné a v reálném čase snímané pozadí. Jak se architekt bude rozhlížet a pohybovat skrze aktuální prostor, bude virtuální verze jeho nápadů interagovat s tím, co on uvidí skrze brýle aktuálně, v reálném čase. Takže se ke zmíněným vymoženostem přidává i detekce pohybu, jak ji známe u automatických polohových systémů. V podstatě tedy jde o složitou percepční, polohovou a pohybovou symbiózu člověka s počítačem.

Druhý příklad. Mobilní robot (říkejme mu/jí „robomobil“) je propojen s člověkem na vzdálenost, a to tak, že může sledovat jeho pohyby (řekněme procházku ulicemi města, bloumání, hru). Dotyčný člověk by asi měl u sebe něco podobného mobilnímu telefonu, pomocí něhož by ho robomobil průběžně lokalizoval a on by obdobně mohl lokalizovat svého robomobila.¹⁴ Nemuseli by se ani ubírat po stejně cestě: roboti chodí přes most, to já můžu louží. Máme tu tedy opět symbiózu, tentokrát „spolupohyb“ člověka a robota – a lze neomezeně teoretizovat o tom, jak mohou spolu objevovat, spoluutvářet a reflektovat prostor k pohybu.

Literatura

- Šikl, R. (2012). *Zrakové vnímání*. Praha: Grada.
- Havel, I.M. (2013). *Zjítréná mysl a kouzelný svět*. Praha: Dokořán.
- Riegl, A. (1893/1985). *Stilfragen. Grundlegungen zu einer Geschichte der Ornamentik*. München: Mäander Verlag GmbH.
- Riegl, A. (1901/1927). *Spätromische Kunstindustrie*. Wien: Druck und Verlag der Osterr. Staats-druckerei.
- Deleuze, G. & Guattari, F. (2010). *Tisíc plošin*. Brno: Hermann a synové.
- Sanouillet, M. & Peterson, E. (Eds.). (1973). *The Writings of Marcel Duchamp*. New York: Oxford University Press.
- Havel, I.M. & Mitášová, M. (2004). Prostor prožívaný jako prostor k jednání. In: *Prostor a jeho člověk*. Praha: Nakl. Vesmír.

Schumacher, P. (1996). The Architecture of Movement. *Arch+*, č. 134/135.

Holl, S. (2003) *Paralaxa*. Brno: Era group.

¹⁴ Něco podobného už exisuje, jmenuje se GITA a lze to použít například jako nosiče zavazadel (dostupné na <www.youtube.com/watch?v=fK8uv0EFaho>).

Vývoj teórie myслe u detí a adolescentov

Martin Jáni

Stredoevropský technologický institut, Masarykova univerzita
Kamenice 5, Brno
martinjani@mail.muni.cz

Abstrakt

Teória myслe je schopnosť porozumieť vlastným mentálnym stavom a mentálnym stavom iných. Objavuje sa počas predškolského veku no jej ďalší vývin do dospelosti je pomerne málo preskúmaný. Dvom skupinám žiakov, 4. stupňa (vek 9-10) a 8. stupňa ZŠ (vek 13-14) bol administrovaný test teórie myслe, ktorý obsahoval 3 krátke príbehy so zložitým sociálnym príbehom. Skupina adolescentov dosiahla lepšie výsledky v teórii myслe aj kontrolných pamäťových otázkach. Pri analýze výsledkov u každého príbehu zvlášť dosiahli v dvoch príbehoch adolescenti v teórii myслe lepšie výsledky než mladšia skupina, výkon v pamäťových otázkach bol porovnatelný. Korelácia medzi kapacitou pamäte a teóriou myслe bola významná len u mladšej skupiny.

1 Úvod

Teória myслe patrí medzi najviac skúmané procesy sociálnej kognície. Označuje schopnosť rozpoznať, predstaviť si a chápať mentálne stavy druhých. Teória myслe, nazývaná aj mentalizácia, je nevyhnutná pre naše každodenné fungovanie. pretože nám pomáha porozumieť myšlienkom, predstavám a presvedčeniam druhých a tým robí našu sociálnu interakciu efektívnejšou (Frith, 2001). Existuje rozsiahla literatúra k vývoju teórie myслe, i keď väčšina sa zameriava na jej vznik a rozvoj v prvých rokoch života, alebo na abnormality u netypickej populácie (Chung, Barch, a Strube, 2014). Niekoľko štúdií ale ukazuje, že vývoj teórie myслe pokračuje ďalej do dospelosti (Happé, Winner, a Brownell, 1998; Kinderman, Dunbar, a Bentall, 1998; Stiller a Dunbar, 2007). Táto schopnosť sa objavuje v predškolskom veku, prakticky všetky zdravé deti dokážu do šiestich rokov pochopiť, že pohľad druhého človeka sa môže lísiť od reality. Zložitejšie testy, ktoré skúmajú online použitie teórie myслe (a teda so zapojením exekutívnych funkcií) odhalujú postupný rozvoj tejto schopnosti počas adolescencie, čo zodpovedá maturácii súvisiacich mozgových štruktúr (mediálny prefrontálny kortex a temporálne oblasti) (Dumontheil, Apperly, a Blakemore, 2010).

1.1 Testy teórie myслe

Väčšina štúdií, ktorá skúmala vývoj teórie myслe, bola zameraná na deti predškolského veku. Existujú rôzne prístupy k meraniu teórie myслe, bolo vyvinutých viacerých testov. Nižšie budú predstavené testy, ktoré sa vzťahujú k vývoju teórie myслe alebo tie, z ktorých čerpá naša behaviorálna paradigma.

1.1.1 Testy nesprávnej domneniek

Najčastejšie používanými sú rôzne varianty testu nesprávnej domneniek, v ktorom má subjekt rozhodnúť, či sa presvedčenie istej osoby zhoduje s realitou (a tým odlísiť od vlastného presvedčenia). Snáď najznámejší je test s bábkami Sally a Ann (Frith, 2001) kde jedna bábka v neprítomnosti druhej schová určitý predmet, po návrate druhej bábky má dieťa určiť, kde bude navrátiť bábku predmet hľadať. Alternatívu predstavuje test s lenticikmi (tzv. Smarties task), kde dieťaťu ukážeme krabičku od lenticiek s farbičkami vo vnútri, následne sa pýtame, čo si bude myslieť jeho kamarát, že je vo vnútri (Perner, Leekam, a Wimmer, 1987). Testy tohto typu sú príliš jednoduché pre staršie deti, pričom môžeme sledovať efekt stropu. Taktiež testy, ktoré sú navrhnuté pre deti predškolského veku, nemusia reflektovať spôsob, akým používajú teóriu myслe staršie deti.

1.1.2 Pokročilé testy teórie myслe

Ako odpoveď na limity testov nesprávnej domneniek boli zhotovené pokročilejšie testy teórie myслe. Happé (1995) vytvorila test pre 8-9 ročné deti, pozostávajúci z príbehov a otázkami pýtajúcimi sa na mentálne stavy protagonistov a nesociálne faktory (kontrolné otázky).

Ďalší pokročilý test, nazvaný Test čítania myšlienok z očí (angl. Reading mind in the eyes test) využíva fotografie oblastí očí. Tie sú prezentované účastníkom, ktorí majú podľa pohľadu očí hádať mentálne stavy (Baron-Cohen, Jolliffe, Mortimore, a Robertson, 1997).

Test zahŕňajúci verbálnu aj neverbálnu modalitu vytvorili Gallagher a spol. (Gallagher a spol., 2000). Verbálnu časť predstavovali texty s príbehmi zatialčo

v neverbálnej časti boli použité krátke príbehy a prezentované v podobe komiksu. Prevažná väčšina splnila tento test, preto boli merané reakčný čas a aktivácia mozgu.

Tieto testy boli vyvinuté k porovnaniu úrovne teórie myслe u ľudí s poruchami autistického spektra a u zdravých.

1.1.3 Škala teórie myслe

Wellman (2001) hovorí, že teória myслe nie je jednoduchá jednodimenzionálna schopnosť, ale skôr komplexný proces, ktorý sa vynára postupne zvládnutím viacerých komponentov zaujímania perspektív dru-hých, ako chápanie zámerov, presvedčení a emócií. K tomu vytvoril škálu teórie myслe (Wellman, Fang, Liu, Zhu, a Liu, 2006; Wellman a Liu, 2004), ktorá pozostáva so siedmich subtestov, lísiacich sa v konceptuálnych rozdieloch a nie v komplexnosti či zapojení iných kognitívnych funkcií.

1.1.4 Zmena perspektív

Deti nadobúdajú schopnosť teórie myслe v pred-školskom veku, zatiaľ čo iné kognitívne domény, ako exekutívne fukcie a s nimi spojené potlačenie egocentrického skreslenia, sa môžu vyvíjať a zrieť neskôr (Dumontheil a spol., 2010). To môžeme overiť pomocou testu s riaditeľom (angl. Director task), kde je úlohou účastníka presúvať predmety uložené na poličke. Niektoré predmety sú z druhej strany zakryté, pričom riaditeľ je v niektorých prípadoch na druhej strane poličky a vtedy zakryté predmety nevidí. Účastník to musí brať do úvahy a počas úlohy tak prepínáť medzi perspektívou riaditeľa a svojou vlastnou, čím my sme mali sledovať interakciu teórie myслe a exekutívnych funkcií (B Keysar, Barr, Balin, a Brauner, 2000; Boaz Keysar, Lin, a Barr, 2003).

1.1.5 Vyšie rády intencionality

V reálnom svete sú sociálne interakcie komplexnejšie, než napr. priama detekcia nesprávneho presvedčenia. Ľudia dokážu pracovať aj s vyššími úrovňami teórie myслe (označovanými ako rády intencionality). Samotná teória myслe predstavuje druhý rám intencionality, teda *ja si myslím* (1. rám), že *ty sa domnievaš* (druhý rám). V reálnych situáciách musíme často napr. sami odhadnúť, čo iná osoba predpokladá, že si my myslíme (3. rám), že chce niekto ďalší (4. rám). Takto by sme mohli teoreticky pokračovať do nekonečna. Kinderman (1998) a spol. ukázali, že ľudia si dokážu predstaviť najviac piaty rám intencionality, potom začínajú zlyhávať. Napriek tomu boli dospelí schopní správne odhadnúť na otázky týkajúce sa kauzálneho reťazca udalostí až do siedmich či ôsmich prvkov.

1.2 Ciel výskumu

Je možné, že kapacita narábať s týmito vyššími rádmi intencionality, podobne ako zapojenie exekutívnych funkcií, sa vyvíja s vekom počas zrenia oblastí mozgu zodpovedných za mentalizáciu, prípadne sa zlepšuje počas neustáleho používania. Máme málo informácií o vývoji teórie myслe po šiestom roku života. Stredný školský vek predstavuje obdobie, v ktorom začínajú deti tráviť viac času so svojimi rovesníkmi, interakcia s deťmi rovnakého veku nadobúda väčšiu dôležitosť. Tento proces kulminuje v adolescencii, predstavuje teda klíčové obdobie z hľadiska uplatnenia sociálnej kognície. Naším cieľom je teda objasniť vývoj teórie myслe v mladšom školskom veku a na začiatku adolescencie.

2 Metóda

2.1 Participanti

Výskumu sa zúčastnili dve skupiny žiakov, 4. a 8. ročník. Žiaci 4. ročníku boli vybraní ako najmladšia kohorta, ktorá má dostatočné schopnosti čítať a porozumieť písanemu textu, pretože už používajú čítanie ako komunikačný nástroj. Žiaci 8. ročníku boli vybraní ako reprezentanti skorej adolescencie, s kognitívnymi schopnosťami (ako kapacita pamäte a logické uvažovanie) na úrovni dospelých. Predpoklad je, že schopnosť mentalizácie sa vyvíja pomalšie a neodpovedá úrovni kapacity pracovnej pamäte. Na druhej strane predpokladáme istú mieru asociácie medzi úrovňou pamäte a mentalizačných schopností.

Zber dát prebehol na Základnej škole na Pražské, Pelhřimov, výskumu sa zúčastnilo 72 žiakov zo štyroch tried, 29 žiakov bolo z 4. ročníku (vek 9-10 rokov) a 43 žiakov z 8. ročníku (vek 13-14 rokov). Účasť na výskume bola dobrovoľná, zákonní zástupcovia žiakov poskytli písomný súhlas. Žiadny zo žiakov nemal poruchu čítania. Po skončení výskumu žiaci uviedli, že sa predtým nezúčastnili podobného výskumu a hypotézy im neboli vopred známe.

2.2 Úloha

K meraniu teórie myслe sme zvolili dizajn, ktorí predtým použili Kinderman a spol. (1998) a Stiller a Dunbar (2007), pričom sme ho prispôsobili veku účastníkov. Žiaci boli testovaní v konferenčnej miestnosti, v jednom čase vždy po jednej triede. Žiaci dostali inštrukcie, následne im boli prezentované príbehy prostredníctvom dataprojektoru. Každý príbeh obsahoval komplexný sociálny scenár. Poradie bolo príbehov bolo prezentované rozdielne pre triedy A a B. Každý príbeh bol prezentovaný istý stanovený čas, ktorý zodpovedal dĺžke príbehu. Po prezentácii príbehu bol ten skrytý z plátna a žiaci mali odpovedať na otázky týkajúce sa konkrétneho príbehu na odpoveďových hárkoch. Odpoveď bola formou nútenej

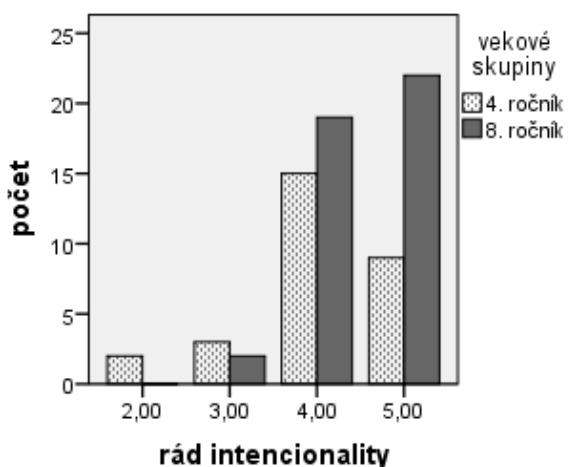
voľby medzi dvoma možnosťami. Žiaci odpovedali na dva druhy otázok. Párne položky sa pýtali na mentálne stavy protagonistov, so stúpajúcimi rádmi intencionality do piatej úrovne. Párne položky mali formu kontrolných otázok zameraných na pamäť, pozornosť a schopnosť čítať, pýtali sa na lineárny reťazec udalostí, so stúpajúcim množstvom prvkov, tiež do piatej úrovne. Za každú správnu odpoveď bol udelený jeden bod. Body z pamäťových položiek boli jednoducho spočítané, zatiaľ čo položky k teórii myсле boli počítané zvlášť pomocou váženého priemeru, pričom výhody boli prisúdené rámom intencionality.

3 Výsledky

Žiaci 8. triedy dosiahli vyššie skóre v položkách teórie myсле ($U = 917$; $p = 0,001$) aj pamäte ($U = 856,5$; $p = 0,004$). Ak sa však pozrieme na každý príbeh zvlášť, zistíme, že len jeden príbeh ukázal rovnaký vzorec. V ďalších dvoch príbehoch boli ôsmaci signifikantne lepší než štvrtáci len v položkách teórie myсле, zatiaľ čo rozdiely v položkách pamäte neboli štatisticky významné (pre podrobnejšie informácie o dosiahnutých výsledkoch v teórii myсле aj pamäti viď Tabuľku 2).

Spearanova neparametrická korelácia medzi úrovňou teórie myсле a pamäťovými schopnosťami ukázala na tesný vzťah ($rs = ,50$; $p < ,001$). Avšak keď sme analyzovali vekové skupiny samostatne, zistili sme tesný vzťah medzi teóriou myсле a pamäťovými schopnosťami len u žiakov 4. (rs = ,60; $p = ,001$), ale nie žiakov 8. ročníku (rs = ,16; $p = ,311$). Celkovo rozptyl v skóre teórie myсле môžeme vysvetliť úrovňou pamäte z 25%, a ak sa budeme dívať len na žiakov

štvrteho, tak najviac z 36% (prehľad korelácií teórie myсле s kapacitou pamäte podľa skupín i celkovo viď Tabuľku 1.).



Obr. 1: Počet žiakov podľa dosiahnutej úrovne intencionality (priemer) a podľa ročníku.

skupina	spearmanovo rho	p	R ²
4. ročník	0,601	0,001	0,36
8. ročník	0,158	0,311	0,03
celkom	0,504	< 0,001	0,25

Tab. 1: Korelácia teórie myсле a kapacity pamäte podľa skupín

Príbeh	Typ položky	4. ročník			8. ročník			U	z	p	r
		Medián	Priemer	s	Medián	Priemer	s				
Celkom	Intencionalita	4,05	3,99	0,79	4,59	4,58	0,48	917,00	3,43	0,001	0,40
	Pamäť	13,00	12,66	1,20	14,00	13,40	0,90	856,50	2,88	0,004	0,34
Príbeh 1	Intencionalita	4,00	3,37	0,94	4,00	3,75	0,65	423,00	2,09	0,036	0,25
	Pamäť	4,00	3,79	0,56	4,00	3,91	0,37	667,00	0,70	0,331	0,11
Príbeh 2	Intencionalita	3,93	3,77	0,94	5,00	4,29	0,96	834,00	2,52	0,012	0,30
	Pamäť	4,00	4,21	0,82	5,00	4,67	0,47	821,00	2,59	0,010	0,30
Príbeh 3	Intencionalita	5,00	4,06	1,23	5,00	4,80	0,49	825,50	2,94	0,003	0,35
	Pamäť	5,00	4,66	0,61	5,00	4,81	0,39	687,50	1,02	0,309	0,12

Tab. 2: Rozdiely medzi skupinami vo výkone teórie myсле a v pamäťovej kapacite

4 Diskusia

Výsledky ukazujú, že teória myсле, rovnako ako kapacita pamäte, sa vyvíjajú od staršieho školského veku do adolescencie. To súhlasí so zisteniami, že

teória myсле sa vyvíja aj po predškolskom veku a aj v dospelosti. Adolescenti dosiahli priemerne na 4,58 rádu intencionality, čo zhruba odpovedá hornému limitu (okolo 5. rádu) zistenému v predošlých výskumoch u dospelých (Kinderman a spol., 1998; Stiller a Dunbar, 2007). Táto štúdia bola dizajnovaná

tak, že maximálna dosiahnutelná úroveň bola 5., dátá teda ukazujú efekt stropu. Je možné, že pri vyšom limite by dátá ukazovali centrálnu tendenciu okolo 5. úrovne. Mladšia skupina dosiahla mierne horší výsledok s rámom intencionality 3,99. To ukazuje, že nedochádza k dramatickej zmene, ale skôr ku kontinuálnemu vývoju. Je teda možné, že takýto rozvoj teórie myсле u 4. ročníku je (a) v dôsledku maturačných procesov v mozgu (Dumonttheil a spol., 2010), alebo (b) v dôsledku vystavenia zvýšenej sociálnej interakcii, príznačnej pre tento vek, ktorá vyžaduje práve intenzívne využívanie teórie myсле a jej precvičovanie (Lewis, Freeman, Kyriakidou, Maridakis-Kassotaki, a Berridge, 1996; Ruffman, Perner, Naito, Parkin, a Clements, 1998).

Kapacitou pamäte bolo možné vysvetliť 25% rozptylu v teórii myсле, u štvrtákov až 36%. Podobne kapacita pamäte vysvetlovala 35% rozptylu v teórii myсле u dospelých (Stiller a Dunbar, 2007), čo odpovedá našej skupine 4. ročníku. Korelácia medzi kapacitou pamäte a teóriou myсле nebola u žiakov 8. ročníku významná. To by mohlo naznačovať, že sociálne ako aj nesociálne fakty sú rovnako zaujímavé pre štvrtákov. Na druhej strane ôsmakov viac zaujímajú sociálne fakty, čo sa odzrkadľuje aj v ich lepšom výkone v teórii myсле. Treba však upozorniť, že i keď boli položky vytvorené tak, aby sa kontrolné otázky pýtali na základné fakty nesúvisiace so sociálnymi aspektami príbehu, tie boli stále súčasťou sociálneho kontextu scenára. Navyše, bolo preukázané, že ľudia si lepšie pamätajú sociálne, než nesociálne fakty (Mesoudi, Whiten, a Dunbar, 2006).

Je tiež možné, že úroveň teórie myсле je u štvrtákov stále závislá na kapacite pamäte, zatiaľ čo u staršej skupiny je krátkodobá pamäť na úrovni dospelých a teda teória myсле je merateľná sama o sebe. Obe skupiny ale dosiahli lepší výkon v pamäťových otázkach, než v otázkach na teóriu myсле. Treba však zdôrazniť, že aj keď obe skupiny položiek boli vytvorené tak, aby boli porovnatelne rozsiahle a obsahovali rovnakú úroveň komplexnosti, predsa len rády intencionality predstavujú iný druh úrovne, oproti pamäťovým položkám (Kinderman a spol., 1998). Nakoniec, žiaci 4. ročníku by už mali vedieť plynule čítať a sústrediť sa na samotnú úlohu, nemôžeme však vylúčiť, že schopnosť čítania (ktorá je závislá na používaní a precvičovaní) sa stále vyvíja.

Neboli zistené žiadne signifikantné rozdiely medzi pohlaviami v teórii myсле či kapacite pamäti, ani v rámci vekových skupín. Štúdie na dospelých jedincoch súčasťou ukazujú na rozdielny výkon v teórii myсле u mužov a žien (Stiller a Dunbar, 2007), obe pohlavia využívajú čas konverzácia približne podobne, sociálne záležitosti zaberajú asi dve tretiny času (Dunbar, 1996). Tieto zistenia naznačujú, že vývoj teórie myсле je porovnatelný u chlapcov aj dievčat.

4.1 Limity

Vzhľadom na efekt stropu si nemôžeme byť istí, či vrchný limit intencionality leží u našich skupín okolo piateho rádu. Bolo by zaujímavé vidieť vrchný limit, ak by experiment zahŕňal dosiahnutelné vyšie rády intencionality. Avšak vzhľadom na najvyššie rády dosiahnuté dospelými v predchádzajúcich študiách nepredpokladáme, že by sa priemerné skóre dramaticky líšilo.

Ďalšie obmedzenie výskumu predstavuje spôsob, akým boli dátá zozbierané. Výber vzorky bol nenáhodný, pozostával z všetkých tried 4. a 8. ročníku jednej základnej školy. Výskumný súbor je relatívne malý, je však tvorený žiakmi menšieho mesta v centrálnej časti Českej republiky, kde značná časť dochádza z okolitých obcí. Môžeme teda predpokladať, že výskumný súbor zahŕňa rozličné socio-ekonomicke skupiny, čím by sa vyvažovalo možné skreslenie jedného socio-ekonomickeho prostredia (Cole a Mitchell, 2000; Cutting a Dunn, 1999). To ale môže prinášať iné obmedzenie: škola, ktorá je stratávacím miestom detí z rôznych prostredí, môže poskytovať rozličné pohľady a tým podporovať rozvoj sociálnej kognície (Lewis a spol., 1996; Ruffman a spol., 1998).

Odhliadnúc od týchto limitácií, štúdia sa zameriava na pomerne nepreskúmanú oblasť. V budúcnosti by bolo prínosné preskúmať vývoj teórie myсле od detstva cez dospelosť až do staroby. S ohľadom k tomu bude potrebné vytvoriť príbehy, ktoré by merali vyššie rády intencionality než piaty. Zároveň, vzhľadom na to, že žiaci 4. triedy si viedli pomerne dobre, bolo by vhodné vytvoriť príbehy pre mladšie deti, aby sme mohli sledovať vývoj teórie myсле od preškolského do školského veku. K tomuto účelu ale bude nutné orientovať sa na iné formy, napr. použitím komiksov alebo videa, aby sme odstránili možné problémy súvisiace s čítaním a pamäťou.

5 Záver

Teória myсле je schopnosť porozumieť mentálnym stavom druhých, objavuje sa počas predškolského veku a vyvíja sa až do neskorej dospelosti. Stredný školský vek je dôležité obdobie pre rozvoj teórie myсле, keďže jedinci v tomto veku začínajú tráviť viac času s rovesníkmi a precvičovať svoje sociálne zručnosti. Tento vývoj kulminuje v adolescencii, kedy interakcia s rovesníkmi zaujme hlavné miesto. Zároveň je to ale pomerne nepreskúmané obdobie vývoja teórie myсле. V tomto výskume sme porovnali vývoj teórie myсле u žiakov 4. a 8. Zistili sme, že sociálne-kognitívne schopnosti pokračujú vo vývoji, pričom výkon adolescentov je už porovnatelný s dospelými. Skupina staršieho školského veku mala oproti adolescentom mierne slabší výkon, pričom tento výkon koreloval s kapacitou pamäte. Rozdiely medzi pohlaviami neboli významné. Vzhľadom na veľkosť vzorku a dizajn výskumu ale musíme brať výsledky ako predbežné, ďalší výskum je

potrebný pre lepšie porozumenie vývoja teórie myслe medzi predškolským vekom a dospelosťou.

Literatúra

- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal of Child Psychology and Psychiatry*, 38(7), 813–822. <http://doi.org/10.1111/j.1469-7610.1997.tb01599.x>
- Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology*, 18(2), 279–295. <http://doi.org/10.1348/026151000165698>
- Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: individual differences and interrelations. *Child Development*, 70(4), 853–865.
- Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331–8. <http://doi.org/10.1111/j.1467-7687.2009.00888.x>
- Dunbar, R. I. M. (Robin I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press. Retrieved from <http://www.hup.harvard.edu/catalog.php?isbn=9780674363366>
- Frith, U. (2001). Mind Blindness and the Brain in Autism. *Neuron*, 32(6), 969–979. [http://doi.org/10.1016/S0896-6273\(01\)00552-9](http://doi.org/10.1016/S0896-6273(01)00552-9)
- Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia*, 38(1), 11–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10617288>
- Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–55. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7789204>
- Happé, F. G. E., Winner, E., & Brownell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology*, 34(2), 358–362. <http://doi.org/10.1037/0012-1649.34.2.358>
- Chung, Y. S., Barch, D., & Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin*, 40(3), 602–16. <http://doi.org/10.1093/schbul/sbt048>
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11228840>
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25–41. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12893123>
- Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2), 191–204. <http://doi.org/10.1111/j.2044-8295.1998.tb02680.x>
- Lewis, C., Freeman, N. H., Kyriakidou, C., Maridakis-Kassotaki, K., & Berridge, D. M. (1996). Social influences on false belief access: specific sibling influences or general apprenticeship? *Child Development*, 67(6), 2930–2947.
- Mesoudi, A., Whiten, A., & Dunbar, R. (2006). A bias for social information in human cultural transmission. *British Journal of Psychology (London, England : 1953)*, 97(Pt 3), 405–423. <http://doi.org/10.1348/00712605X85871>
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5(2), 125–137. <http://doi.org/10.1111/j.2044-835X.1987.tb01048.x>
- Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, 34(1), 161–174.
- Stiller, J., & Dunbar, R. I. M. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1), 93–104. <http://doi.org/10.1016/j.socnet.2006.04.001>
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, 72(3), 655–684. <http://doi.org/10.1111/1467-8624.00304>
- Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. *Psychological Science*, 17(12), 1075–1081. <http://doi.org/10.1111/j.1467-9280.2006.01830.x>
- Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, 75(2), 523–541. <http://doi.org/10.1111/j.1467-8624.2004.00691.x>

Computational models of homeostatic synaptic and dendritic plasticity in the hippocampus

Peter Jedlička

Institute of Clinical Neuroanatomy, Neuroscience Center, Goethe University Frankfurt, Frankfurt/Main, Germany
E-mail: jedlicka@em.uni-frankfurt.de

Abstract

I present our recent work on homeostatic aspects of (1) synaptic and (2) dendritic plasticity: (1) We used biologically realistic models of dentate granule cells to simulate LTP and heterosynaptic LTD observed in the hippocampal dentate gyrus. LTP and LTD could be accounted for by a combination of a spike-timing-dependent plasticity (STDP) mechanism and a fast homeostatic Bienenstock-Cooper-Munro (BCM)-like metaplasticity mechanism. (2) We used biologically realistic models of dentate granule cells to simulate functional consequences of denervation-induced dendritic plasticity. Our results showed that dendritic retraction boosted the excitability of denervated neurons and their responsiveness to synaptic stimulation, contributing to firing rate homeostasis.¹

1 Introduction

Long-term strengthening or potentiation (LTP) and long-term weakening or depression (LTD) of synapses are key mechanisms of learning and memory. Hippocampal dentate gyrus plays an important role in spatial learning. In our recent work we focused on homeostatic aspects of (1) synaptic and (2) dendritic plasticity in the dentate gyrus:

(1) Synaptic plasticity in the form of long-term potentiation (LTP) and long-term depression (LTD) is widely accepted to underly synaptic changes involved in the storage and processing of information in the hippocampus. It remains uncertain, however, which particular activity rules are utilized by hippocampal neurons to induce LTP and LTD in behaving animals. Recent experiments in the dentate gyrus of freely moving rats (Bowden et al., 2012) revealed an unexpected pattern of LTP and LTD from high-frequency perforant path stimulation. While 400 Hz theta-burst stimulation (400-TBS) and 400 Hz delta-burst stimulation (400-DBS) elicited substantial LTP of the tetanized medial path input and, concurrently, LTD of the non-tetanized lateral path input, 100 Hz theta-

burst stimulation (100-TBS, a normally efficient LTP protocol for *in vitro* preparations) produced only weak LTP and concurrent LTD. The results show that a simple STDP plasticity rule with fast BCM homeostasis / metaplasticity can reproduce the Bowden et al. pattern of results when implemented in a compartmental granule cell model with realistic biophysics. The model thus gives insight into the computations that granule cells are making, as driven by plasticity-inducing synaptic events arising from both tetanization and background activity (see below).

(2) In contrast to synaptic plasticity, structural plasticity of dendrites is a much less studied and understood form of neuronal plasticity. Therefore, in the second part of my presentation, I describe anatomically detailed models of hippocampal neurons, which predict that dendritic plasticity is able to homeostatically adjust neuronal excitability and possibly modulate local synaptic plasticity. In our modeling we focused on denervation-induced dendritic plasticity. The phenomenon of denervation-induced dendritic plasticity has been studied in some detail using the classical “entorhinal cortex lesion (ECL) model” (see e.g. Deller & Frotscher, 1997). In this experimental setting entorhinal afferents to dentate granule cells are lost and granule cells profoundly remodel their dendritic tree (e.g. Vuksic et al., 2011). Entorhinal denervation causes a loss of dendrites of dentate granule cells resulting in less complex dendritic arbors and a persistent shortening of the dendritic tree. The functional consequences of this shortening or retraction, which typically have been regarded as “atrophic”, damaging or detrimental to the denervated neuron, have not been studied and remained elusive.

(1) & (2): In the discussion I suggest that future studies should be exploring how synaptic and dendritic plasticity interact with each other.

2 Results

(1) To simulate heterosynaptic plasticity, we adopted a published, multicompartmental and biophysically realistic computational NEURON model of a dentate gyrus granule cell (Aradi and Holmes, 1999; Santhakumar et al., 2005). This nine-section, 125 compartment granule cell model received input from 150 medial path synapses and 150 lateral path

¹ Parts of this presentation have been previously published by PLOS Computational Biology and Acta Neuropathologica Communications in open access articles (Jedlicka et al., 2015; Platschek et al., 2016) under the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction.

synapses, distributed in appropriate zones across the two dendritic branches (Fig. 1).

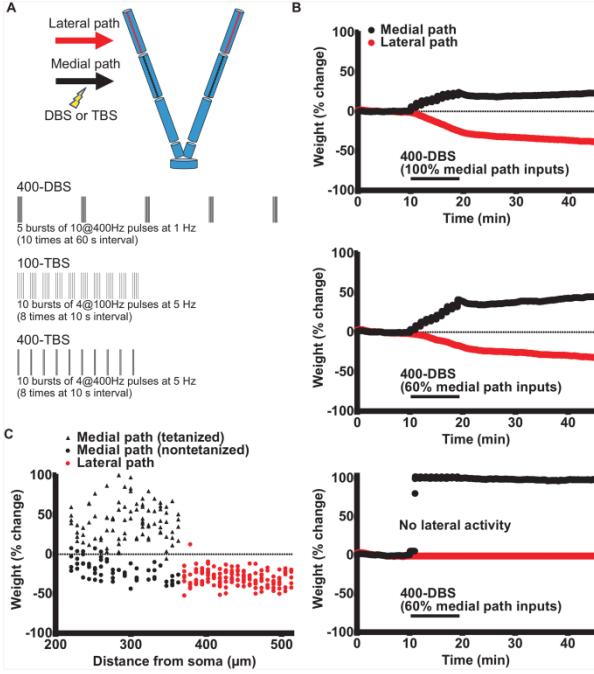


Fig. 1: Biophysically realistic granule cell model for generating LTP on the medial path and concurrent LTD on the lateral path (Jedlicka et al., 2015).

The schematic in Fig. 1A shows the compartmental model. Red and black dots on the dendrites represent positions for medial and lateral perforant path synapses, respectively. Three HFS protocols applied to the medial path were used to test for LTP on the medial path and concurrent LTD on the lateral path: 400-DBS, 100-TBS and 400-TBS (Fig. 1A, bottom). Panel B in Fig. 1 shows perforant path synaptic weights on a simulated granule cell before, during and after 400-DBS of the medial perforant path. Top: 400-DBS delivered to all medial path synapses produced smaller LTP on the medial path than concurrent LTD on the lateral path. Middle: 400-DBS delivered to 60% of medial path synapses produced greater LTP than concurrent LTD. Bottom: There was a loss of lateral path LTD when ongoing activity in that pathway was set to zero after the onset of 400-DBS (no lateral activity), demonstrating the need for ongoing spontaneous synaptic activity to drive LTD in the non-tetanized synapses. All graphs depict average values for all given synapses over 3 runs. Panel C in Fig. 1 shows spatial distribution of synaptic weight changes for 400-DBS delivered to 60% of medial path synapses, showing that non-tetanized medial synapses exhibit LTD. Weight changes are expressed as % change with respect to their baseline value.

The model was endowed with a synaptic plasticity mechanism containing a presynaptically centered nearest-neighbor implementation of the STDP

rule with fast homeostasis (metaplasticity). The multicompartmental granule cell model included the biophysics of the main ion channels in the dendrites and soma. Action potentials were generated in the soma and back-propagated along the dendrites, both electrotonically as well as due to the action of dendritic sodium and calcium ion channels. Thus the granule cell model took into account all of the complex spatio-temporal integration of EPSPs in the dendrites that are evoked by the spontaneous activity, the experimental high-frequency stimulation (HFS) protocol, and the back-propagating postsynaptic spikes. In addition, we employed a realistic simulation of the granule cell spontaneous input activity according to in vivo data that show a significant peak around 8 Hz for both the medial and lateral pathways.

Figure 2 summarizes the plasticity results generated by the compartmental model, and compares them with the experimental data obtained by Bowden et al. (2012).

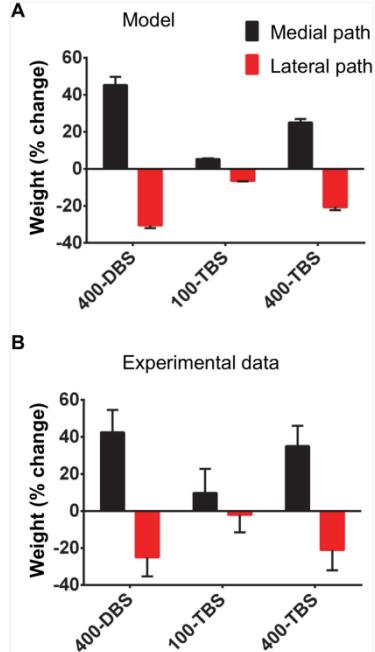


Fig. 2: Summary of effects of stimulation pattern on LTP and concurrent LTD in compartmental model simulations and experiments (Jedlicka et al., 2015).

400-DBS and 400-TBS produced greater LTP and concurrent LTD than 100-TBS in silico for the optimal values of parameters (average results from 3 runs, Fig. 2A) in line with in vivo data (Bowden et al., 2012) which are depicted in Fig. 2B.

The multicompartmental model of granule cell, with realistic active and passive properties, was able to reproduce different experimental results arising from the various experimental protocols (for details see Jedlicka, Benuskova, & Abraham, 2015). The main findings of our modeling study were:

(a) Combined STDP and BCM rules can reproduce the LTP and heterosynaptic LTD, (b) as long as spontaneous activity continues in the input pathways. (c) The degree of LTD depends on the degree of LTP, due to the implemented homeostatic BCM rule that stabilizes cell firing rate. (d) Standard 100 Hz-TBS gives counter-intuitively poor LTP and LTD because this protocol is very good at firing granule cells, which in turn causes the potentiation amplitude parameter to transiently decline, hence braking LTP.

(b) To model denervated-induced dendritic plasticity, we combined morphological and compartmental modeling. By combining full 3D-reconstructions (Vuksic et al., 2011) with published values of passive membrane and cytoplasmic parameters (Schmidt-Hieber et al., 2007), we obtained compartmental models for control (healthy) and denervated dentate gyrus granule cells (Figure 3A). Interestingly, the efficacy of action potential backpropagation was higher in denervated than in control granule cells (Figure 3B). Most strikingly, significant increase in maximum backpropagating action potential (bAP) amplitudes was observed selectively in the denervated dendritic layer (Figure 3B). This result suggests that lesion-induced changes in dendritic morphology facilitate the invasion of bAPs selectively into the deafferented dendritic region.

Fig. 3A shows representative examples for 3-D reconstructions of control (black) and denervated (red) GFP-positive granule cells (Vuksic et al., 2011). Dendograms of each tree are shown to the right. In Fig. 3B simulated action potential backpropagation (bAP) is plotted for control (black) and denervated (red) granule cells ($n=15$, each) as a function of depth in the molecular layer. The border between inner and denervated outer molecular layer is indicated by a dashed line. The IML/OML layer boundary was obtained from the reconstructed cell data.

To predict the consequences of dendritic atrophy (retraction) for the granule cell input-output function we added to the passive model active channels for generating realistic spiking (Schmidt-Hieber and Bischofberger, 2010). As predicted by the higher input resistances in compartmental models of denervated granule cells, the somatic f-I curves were shifted, rendering the neurons more excitable (Fig. 3C, D). The mean output frequencies in somatic f-I curves were strongly increased in denervated cells as compared to control cells. However, when we distributed synapses in dendrites at the same density for compartmental models of both denervated and control cells, we observed similar firing rates (Fig. 3E, F). Thus, in synaptic f-I curves, the greater excitability effectively compensated for the smaller actual number of synapses in the shorter dendrites. Importantly, such remarkable homeostatic compensation for the lower number of synaptic inputs was already present in the passive model, leading to similar somatic voltage output in

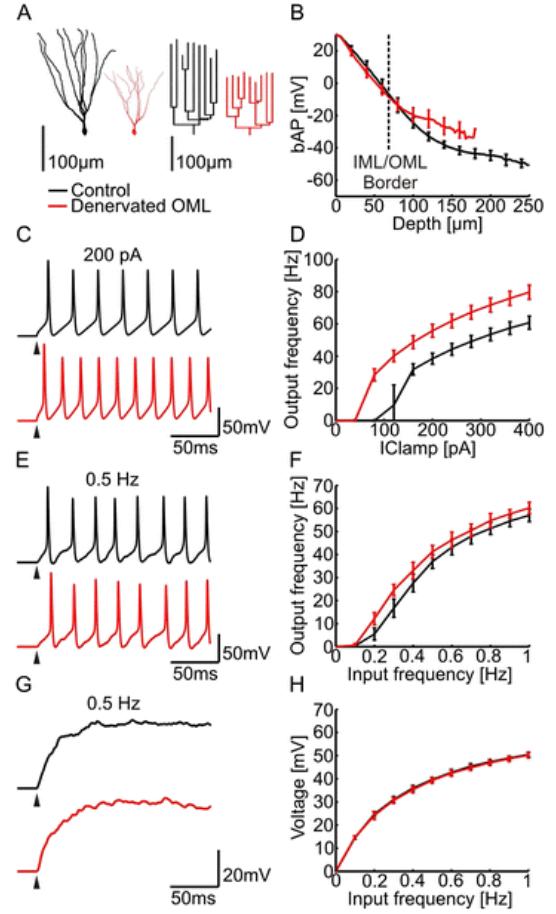


Fig. 3: Action potential backpropagation and excitability in compartmental models of denervated granule cell dendrites (Platschek et al., 2016).

control and denervated cells when all dendritic synapses, distributed again at the same density, were activated (Figure 3G, H). Viewed together, these simulations demonstrate that dendritic remodeling following entorhinal denervation enhances the firing ability of dentate gyrus granule cells and contributes to a homeostatic regulation of their synaptically driven output.

Exploring dendritic reorganization in a variety of morphological models, we found that both phenomena (boost of bAPs and homeostatic enhancement of excitability) that we described for the dentate gyrus granule cell are applicable to any synthetic dendritic tree that we generated. E.g. the homeostatic regulation of excitability was also explained using a simple morphological model: We increased the length of synthetic dendrites grown in a square area by increasing the complexity (Figure 4A). The input conductance increased linearly with the length of dendrite (Figure 4B). Since the average diameter was the same in all cases and the input conductance increases with membrane surface this is not surprising. Because the synapse density does not change, the number of synapses also grows linearly with the length of dendrite. These two measures cancel each other to

reveal an almost constant somatic membrane potential deflection when all synapses are stochastically activated at the same rates (Figure 4C). For the granule cell synthetic and reconstructed morphologies used in our study the linear relation between total dendrite length and input conductance was true (Figure 4D and F), which led to an excitability of the cells that was independent of total length (Fig. 4E and G).

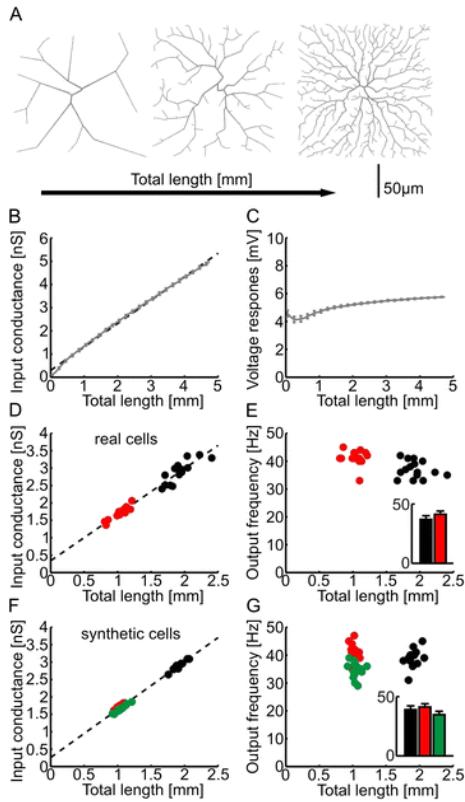


Fig. 4: Morphological models reveal the general electrotonic principle leading to the homeostatic regulation of excitability (Platschek et al., 2016).

3 Discussion

We have found that combining STDP and BCM-like rules with spontaneous activity can replicate the outcomes of three separate *in vivo* experiments in a biologically realistic granule cell model. The vast majority of computational studies that model synaptic plasticity neglect the fact that *in vivo* neurons exhibit an ongoing spontaneous spiking which affects the dynamics of synaptic changes. We have studied how key components of learning mechanisms in the brain, namely spike timing-dependent plasticity and metaplasticity, interact with spontaneous activity in the input pathways of the neuron. Using biologically realistic simulations we have shown that ongoing background activity is a key determinant of the degree of long-term potentiation and long-term depression of synaptic transmission between nerve cells in the hippocampus of freely moving animals. This work helps better understand the computational rules which drive synaptic plasticity *in vivo*. The ongoing

spontaneous input activity can explain why the heterosynaptic LTD is routinely seen in the dentate gyrus *in vivo*, when the hippocampal circuitry is intact while absent in the dentate gyrus *in vitro*, when the input is severed.

Metaplasticity is defined as the activity dependent and persistent change in neuronal state that shapes the direction, duration and/or magnitude of future synaptic change (Abraham, 2008). A typical computational implementation of synaptic metaplasticity theory is the BCM-like model of synaptic plasticity in which the sign and magnitude of plasticity, as well as the position of the sliding modification threshold, are governed by the level of postsynaptic activity averaged over some past. The BCM-like sliding modification threshold serves a homeostatic function by producing cell-wide changes that keep synaptic plasticity within a working dynamic range and flexible. This has the net effect of keeping both LTP and LTD readily available to respond to future changes in correlated presynaptic and postsynaptic activity. In our model, we implemented BCM-like metaplasticity, in which the metaplastic state affects all synapses across the cell. A critical parameter of the model that affects BCM calculations is the length of the cell-firing integration period. Our model was the most stable and robust when using a relatively fast integration period of one minute. This result is consistent with recent network modeling studies (Zenke et al., 2013, 2015).

In the study focusing on dendritic plasticity, we used computational models to predict the effects of dendritic plasticity on the electrotonic structure and intrinsic excitability of denervated dentate granule cells. In the specific case of the denervated granule cells, our simulations revealed that these neurons are, as expected, electrotonically more compact than control granule cells. As a consequence, backpropagating action potential (bAP) attenuation was significantly reduced in their denervated dendrites. Notably, this boost of bAPs was restricted to the denervated dendritic layers. Since bAPs that reach back to the synapse are thought to be involved with strengthening synaptic weights (i.e. with synaptic plasticity), this result indicates that dendritic remodeling could contribute to a homeostatic strengthening of surviving synapses on denervated dendritic segments. Moreover, simulations of somatic and dendritic f-I curves revealed an increased excitability of the cells. In conjunction with the smaller number of synapses this led to a homeostatic maintenance of firing rates in denervated granule cells. Thus, the shortening of dendrites helps to restore the normal firing pattern of granule cells and normalizes information throughput to the hippocampus.

Importantly, using generalized morphological models we unraveled the general principles that led to these two homeostatic features of dendritic remodeling:

a) Using a general morphological model, we were able to show that the spatially selective enhancement of action potential backpropagation is not only present in

the specific case of denervated granule cells but, remarkably, in any dendritic morphology. The retraction of a single dendritic branch or a large dendritic region will lead to a tightly focused reduction in the attenuation of bAPs in the targeted area. This mechanism will enable any dendritic tree which undergoes structural remodeling including extension or shortening of its distal branches to adjust local synaptic plasticity specifically in the remodeled dendritic subcompartments.

b) Remarkably, the increase in excitability of granule cells after denervation compensated exactly for the smaller number of excitatory synapses, suggesting that these structural changes of the granule cell dendritic tree return granule cell excitability to a physiological working level. In other words, by remodeling the size of its dendritic tree a denervated neuron can exquisitely calibrate its excitability and adapt it to the available afferent input. Such a mechanism counteracts the denervation effects and appears to be a homeostatic mechanism by which neurons regulate their activity under injury conditions. Similarly to the bAP enhancement, using a general morphological model, we have shown that this homeostatic mechanism is not only present in granule cells but represents a fundamental feature of all dendritic trees which undergo lengthening or shortening of their branches while keeping synaptic density constant. A change in input conductance gets canceled out by a change in the number of synapses leading to firing rate independent of the length of dendrites.

Collectively, our data imply that denervation-induced structural adaptations of neurons counteract the loss of synaptic inputs due to denervation and thus contribute to neuronal homeostasis. In contrast to the current view, which regards denervation-induced dendritic retraction as detrimental for a neuron, we have proposed that this form of dendritic remodeling returns a denervated neuron to its functional state and may, in fact, be restorative.

(1) & (2): Synaptic plasticity is dependent on postsynaptic depolarization, which may arise from activation of synapses but may also be enhanced by bAPs. Since dendritic plasticity affects both action potential backpropagation as well as neuronal excitability, it can potentially influence synaptic plasticity. However, the link between synaptic and dendritic plasticity has not yet been explored. Therefore, future experimental and computational studies could provide new insights into the interaction between these two important forms of neuronal plasticity.

Acknowledgements

The work leading to presented results was supported by a Young Investigators Grant (from the faculty of medicine Goethe-University), BMBF (No. 01GQ1203A, No. 01GQ1406), Deutsche

Forschungsgemeinschaft (CRC 1080), Croatian Science Foundation (No. 7379).

References

- Abraham WC (2008) Metaplasticity: tuning synapses and networks for plasticity. *Nat Rev Neurosci* 9:387.
- Aradi I, Holmes WR (1999) Role of multiple calcium and calcium-dependent conductances in regulation of hippocampal dentate granule cell excitability. *J Comput Neurosci* 6:215–235.
- Bowden JB, Abraham WC, Harris KM (2012) Differential effects of strain, circadian cycle, and stimulation pattern on LTP and concurrent LTD in the dentate gyrus of freely moving rats. *Hippocampus* 22:1363–1370.
- Deller T, Frotscher M (1997) Lesion-induced plasticity of central neurons: sprouting of single fibres in the rat hippocampus after unilateral entorhinal cortex lesion. *Prog Neurobiol* 53:687–727.
- Jedlicka P, Benuskova L, Abraham WC (2015) A Voltage-Based STDP Rule Combined with Fast BCM-Like Metaplasticity Accounts for LTP and Concurrent “Heterosynaptic” LTD in the Dentate Gyrus In Vivo. *PLoS Comput Biol* 11:e1004588.
- Platschek S, Cuntz H, Vuksic M, Deller T, Jedlicka P (2016) A general homeostatic principle following lesion induced dendritic remodeling. *Acta Neuropathol Commun* 4:19.
- Santhakumar V, Aradi I, Soltesz I (2005) Role of mossy fiber sprouting and mossy cell loss in hyperexcitability: a network model of the dentate gyrus incorporating cell types and axonal topography. *J Neurophysiol* 93:437–453.
- Schmidt-Hieber C, Bischofberger J (2010) Fast sodium channel gating supports localized and efficient axonal action potential initiation. *J Neurosci* 30:10233–10242.
- Schmidt-Hieber C, Jonas P, Bischofberger J (2007) Subthreshold dendritic signal processing and coincidence detection in dentate gyrus granule cells. *J Neurosci* 27:8430–8441.
- Vuksic M, Del Turco D, Vlachos A, Schuldt G, Müller CM, Schneider G, Deller T (2011) Unilateral entorhinal denervation leads to long-lasting dendritic alterations of mouse hippocampal granule cells. *Exp Neurol* 230:176–185.
- Zenke F, Agnes EJ, Gerstner W (2015) Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nat Commun* 6:6922.
- Zenke F, Hennequin G, Gerstner W (2013) Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS Comput Biol* 9:e1003330.

Primitiva virtuální interakce: základní stavební kameny pro vývoj a výzkum interaktivních virtuálních prostředí

Vojtěch Juřík, Pavel Ugwitz, Čeněk Šašinka

Psychologický ústav Filozofické fakulty MU, HUME Lab
Arna Nováka 1, 602 00, Brno

jurik.vojtech@mail.muni.cz, 172577@mail.muni.cz, cenek.sasinka@mail.muni.cz

Abstrakt

Interaktivní virtuální prostředí představují přirozený posun ve vývoji externích informačních reprezentací — vizualizaci. Nejčastější a nejrealističtější formou reprezentace prostoru ve virtuálním prostředí bývají geovizualizace, potažmo jejich deriváty ve formě map. Tyto kartografické produkty jsou v interaktivní formě využívány nejen v mnoha aplikovaných oblastech, ale také v základním výzkumu jako vhodný materiál pro zkoumání řady kognitivních procesů a mentálních kompetencí. V současné době neexistuje jednotná taxonomie pro pojmenování a výzkum základních prvků, které tvoří interaktivní virtuální prostředí. Interakční primitiva představují základní stavební jednotky, kterými lze vyjádřit povahu dynamické interakce s virtuálními prostředími. S mohutným rozvojem různých forem interaktivních vizualizací je proto nezbytné definovat empiricky zakotvenou a ekologicky validní taxonomii interakčních primitiv, která definují povahu celé interakce. V rámci tohoto příspěvku shrnujeme dosavadní poznatky o interakčních primitivech, která představujeme v kontextu některých modelů shrnujících interakci člověka s na mapách založenými vizualizacemi a implikujeme jejich užití při vývoji, výzkumu a aplikaci virtuálních interaktivních vizualizací.

1 Informační vizualizace ve VR

Reprezentace informací nabývá různých forem. S rozvojem informačních technologií se otevírají nové, dosud neprozkoumané možnosti toho, jak lze informace reprezentovat s ohledem na jejich žádoucí zpracování člověkem. Vizualizace (nejen virtuální) představuje grafickou formu externí reprezentace informace a je chápána jako efektivní forma prezentování informace (Ware, 2004). Data by měla být zobrazena v co možná nejpochopitelnější podobě (Card, Mackinlay a Schneiderman, 1999). Uvahy o vhodných typech vizualizace nabývají na významu tehdyn, uvědomíme-li si, že typ dané vizualizace výrazně ovlivňuje (pozitivně i negativně) např. proces řešení problémů, jak poukazují už starší studie (např. Bauer a Johnson-Laird,

1993). S rozvojem moderních technologií a jejich výpočetní kapacity je stále více využívána virtuální realita jako prostředek pro efektivní simulování různých prostředí. Virtuální realita umožňuje snadné uzpůsobení konkrétního obsahu virtuálního prostředí a jeho optimalizaci pro rozličné účely, na druhé straně ale nabízí také skvělé možnosti měření behaviorální aktivity člověka, který ve VR pracuje. Nejčastější reprezentací prostoru ve virtuální realitě je geovizualizace, která nabízí komplexní informaci o zobrazovaném terénu. V podstatě se jedná o virtuální mapu.

Mapa je ze své podstaty považována jako interaktivní nástroj (Roth, 2012), jelikož uživatel může s mapou interagovat mnoha způsoby, tedy přizpůsobovat si ji natočením, přehnutím, přiblížením, oddálením, případně do ní vepisovat (Wallace, 2011). Nicméně, nikdy dříve nebyly požadavky interaktivity tak zásadní pro design map, jako v dnešní době (Andrienko a Andrienko, 1999; Dykes, 2005; Harrower, 2008), protože výrobky moderních technologií umožňují převést mapy do virtuálního prostředí ve formě geovizualizací, které co možná nejlépe odpovídají požadavkům potenciálního uživatele. Virtuální geovizualizace nabízí např. možnosti mapou řízeného usuzování (*map-driven reasoning*) v reálném čase (MacEachren a Monmonier, 1992) a celou řadu dalších funkcionalit a variant.

Formy geovizualizací se různí v jejich zpracování, podobě a obsahu, ale také ve formě, v jaké se nabízí interakce s nimi. Geovizualizace mohou nabývat buď statické nebo interaktivní formy při reprezentování terénu. Ačkoliv dvě mapy mohou být informačně zcela ekvivalentní (Larkin a Simon, 1987), mohou se různit ve způsobech, jakými s nimi lze manipulovat, např. ovládat je, natáčet, přiblížovat, ale třeba také v tom, jaké další informace je možné si určitou akcí zviditelnit. Tento parametr by se dal nazvat interakční ekvalence a představuje srovnatelnost různých vizualizací v možnostech interakce s nimi. Interakční neekvalence dvou geovizualizací (např. statické oproti interaktivní) bude mít za následek odlišné zpracování zobrazeného obsahu na základě odlišných úkonů, které uživatel s geovizualizací provede. Může se stát, že i dvě informačně a gra-

ficky analogické geovizualizace lišící se pouze ve své funkcionalitě (tedy budou interakčně neekvivalentní) povedou ke komputační neekvivalenci (Larkin a Simon, 1987). Z hlediska lidské psychiky je právě povaha specifické geovizualizace zásadní téma, protože nepochopení nebo zkreslení zobrazené informace může představovat bezpečnostní hrozby v aplikovaných oblastech (viz např. Rierson, 2013). Konkrétní povaha dané vizualizace by měla být při tvorbě vždy zohledněna, protože jako taková může významně ovlivnit kognitivní zpracování zobrazované informace (Ware, 2012; Bauer a Johnson-Laird, 1993). Právě povaha budoucí interakce s virtuálním prostředím může být vyjádřena s využitím taxonomie tzv. interakčních primitiv (Roth, 2012), což jsou v podstatě prvky popisující účel interakce s vizualizací (viz níže).

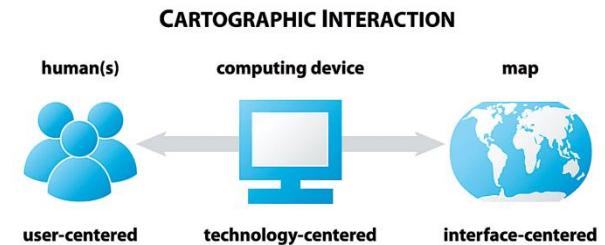
2 Kartografická interakce

Návrhy pro utvoření a ukotvení tzv. interakční vědy byly formovány na základě poznatků z oborů jako informační vizualizace a vizuální analýza, nicméně také na poznatkách a potřebách kartografie a GIScience aj. Někteří autoři zdůrazňují, že informační vizualizace jako taková se skládá ze dvou základních prvků, a to (1) ze samotné reprezentace, ale právě také z (2) interakce, resp. interaktivnosti (Buja a kol. 1996; Kang a kol., 2007; Roth, 2011). Také při tvorbě geovizualizací stojí na jedné straně kartografická reprezentace a na straně druhé kartografická interaktivita. Kartografická reprezentace představuje grafické, zvukové, haptické a jiné parametry mapy ztělesňující samotnou geografickou informaci (Roth, 2012). Tato část podléhala dlouholetému výzkumu mapového zobrazení, především v oblasti percepce (jak člověk mapu vidí), kognice (jak ji rozumí) a sémiotiky (co mapová informace pro člověka znamená) (MacEachren, 1995). Na druhé straně stojí kartografická interaktivita (*cartographic interactivity*), která představuje dialog mezi uživatelem a mapou (Roth, 2011). Tento dialog je zajištěn prostřednictvím konkrétní technologie (rozhraní). Edsall (2003) dále definoval interakční výměnu (*interaction exchange*) coby jednu dílčí sekvenci definovanou v rámci konverzace uživatele s mapou (sekvence „otázka a odpověď“). Celý proces konverzace potom nazývá interakční relace (*interaction session*).

Uživatel a informační vizualizace jsou činitelé interakce a navzájem se ovlivňují. Když uživatel dosahuje prostřednictvím daného rozhraní další informaci vedoucí k jeho cíli, mění si formu dosavadního náhledu na tuto vizualizaci (rovina percepce). Tato změna posléze vede ke změně dosavadní interpretace vizualizace uživatelem (dojde ke změně kognitivního schématu uživatele) a pokud je tato změna žádoucí, uživatel dosáhnul vytýčeného cíle. Pokud ne, je třeba další manipulace s informační vizualizací. Komputační zařízení představuje mediátor dialogu mezi uživatelem a vizualizací, umožňující dynamickou výměnu informací v reálném čase (MacEachren a Monmonier, 1992).

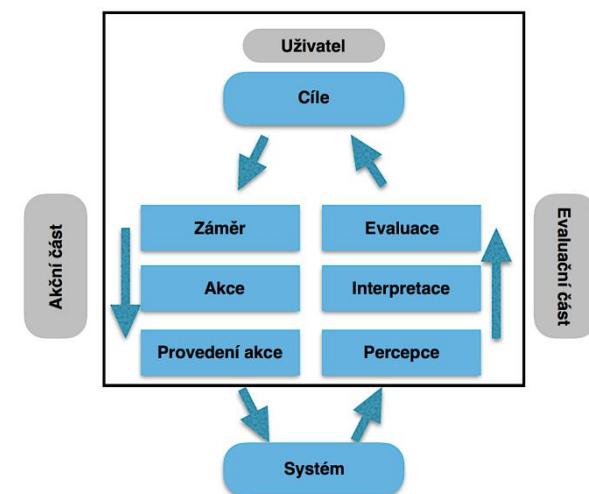
Tato povaha interakce pramení z konceptu konverzační metafory akcentované především v přístupech interakce člověk-počítač, informační vizualizace, a vizuální analýzy (Yi a kol., 2007).

Obecně lze proces kartografické interakce definovat třemi stranami; (1) uživatelem (user-centered perspective), (2) mapou (interface-centered perspective) a (3) výpočetním zařízením (technology-centered perspective), viz obr. 1.



Obr. 1: Schéma kartografické interakce (dle Roth, 2012)

Přesněji je možné interakci člověka s určitým rozhraním demonstrovat s pomocí Normanova sedmi-úrovnového modelu akce (The Seven Stages of Action; Norman, 1988). Model reflekтуje povahu obecné interakce člověka se zařízením a v tomto rámci budou dále diskutována interakční primitiva. Na jedné straně vede uživateli při interakci s vizualizací konkrétní cíl, který se zformuje v záměr vedoucí k akci. Tato akce může mít určitý efekt, který uživatel zpětně reflekтуje, interpretuje a evaluuje, což dále modifikuje jeho původní cíle (viz obr. 2).

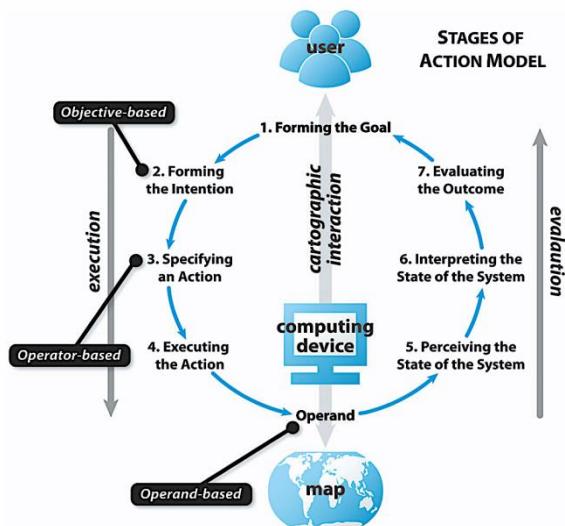


Obr. 2: Normanův sedmi-úrovňový model akce (Norman, 1988)

Interakční primitiva mohou být nejlépe ilustrována v rámci tohoto modelu, pro který přimitiva představují základní jednotky interaktivnosti. Tyto základní jednotky interaktivnosti společně formují interakční výměnu mezi uživatelem a vizualizací (Roth, 2012). Jako taková tedy mohou být mezi sebou interak-

ční primitiva kombinována při použití interaktivních geovizualizací. Přesné definování toho, co interakční primitiva představují, je důležité především z didaktického hlediska, protože tvoří konzistentní taxonomii pro deskripci a vývoj interaktivních map, pro klasifikaci výzkumu interaktivních map a pro jejich zpětné vyhodnocení a ověření funkčnosti.

Ačkoliv jsou interakční primitiva definována primárně pro oblast informační vizualizace, vizuální analýzy, pole human-computer interaction, právě u map je jejich aplikační potenciál zásadní. Systémů popisujících interaktivní primitiva v rámci výše uvedených oblastí existuje v současnosti celá řada (Wehrend a kol., 1990; 1993; Zhou a Feiner, 1998; Blok a kol., 1999; MacEachren a kol., 1999; Crampton, 2002; Andrienko a kol., 2003; Amar a kol., 2005; Yi a kol., 2007) a ačkoliv se jejich definice v mnohem překrývají, neexistuje jejich jednotná taxonomie. Pokusy o jednotící perspektivu uskutečnil Roth (2012; 2013), který vysvětluje interakční primitiva z hlediska různého dělení a hledá jejich hlavní podobnosti a rozdílnosti. Roth (2012) elegantně vykládá interakční primitiva v rámci Normanova modelu (viz výše). Jednotlivá stadia modelu zastupují způsoby, jakým uživatel interaguje s fyzickými, popř. virtuálními objekty (tedy tzv. operandy) a to na úrovni akcí, které jsou ztělesněny konkrétními interakčními primitivy (viz obr. 3). Na každé úrovni tedy hovoříme o sérii různých operací, které uživatel v rámci interakce produkuje.



Obr. 3: Normanův sedmi-úrovňový model zahrnující interakční primitiva (dle Roth, 2012)

3 Interakční primitiva

Roth (2012) dělí interakční primitiva do tří kategorií, podle jejich možného zařazení v rámci Normanova modelu. Na každé úrovni primitiva kategorizují úkony, které jsou v daném stadiu vykonávány uživatelem.

3.1 Primitiva na objektivní bázi

Primitiva založená na objektivní bázi (Objective-based primitives) kategorizují interakci na úrovni formování intence (*Forming the Intention level*, viz obr. 3) a jedná se o operace, které definují konkrétní cíl uživatele při interakci s daným rozhraním. Autori se v jejich výčtu různí (viz Tabulka 1), nicméně tato primitiva se dají obecně chápat jako obecné zadání konkrétního tasku. Pro ilustraci uvádíme vybrané taxonomie autorů, jejichž práce souvisí především s tvorbou virtuálních geovizualizací. Roth (2012) konsoliduje, že i navzdory výrazné variabilitě v uvedené taxonomii, primitiva významu „identifikuj“ (*identify, explore, examine*) a „porovnej“ (tedy vyhledej nebo nalezni podobnosti a odlišnosti, např. *compare*) jsou konzistentně zastoupeny napříč různým dělením (viz tab. 1).

Interakční primitiva	
Blok et al. (1999)	(1) identifikuj, (2) porovnej
Crampton (2002)	(1) prozkoumej, (2) porovnej, (3) (se)řad/(roz)tříd', (4) zdůrazni/upozad', (5) příčina/následek
Yi et al. (2007)	(1) vyber, (2) prozkoumej, (3) přenastav, (4) přířad', (5) abstrahuj/konkretizuj, (6) filtrov, (7) propoj
Wehrend (1993)	(1) identifikuj, (2) lokalizuj, (3) rozliš, (4) kategorizuj, (5) klastruj, (6) seřad', (7) porovnej, (8) spoj, (9) koreluj

Obr. 4: Primitiva založená na objektivní bázi (dle Roth, 2012)

3.2 Primitiva na bázi operátora

Primitiva založená na operátoru (Operator-based primitives) zastupují interakce, které probíhají na úrovni specifikace akce (*Specifying an Action level*, viz obr. 3). Jedná se o primitiva, která souvisí s možností manipulace dané vizualizace. Tato primitiva představují nabídku možných úkonů, na základě kterých může uživatel uskutečnit na cíl zaměřené jednání, nicméně se nejedná o samotné provedení těchto úkonů. V podstatě se jedná o nastavení parametrů uživatelského rozhraní, které vytváří rámec pro možnou interakci. Roth (2012) nachází napříč různými lexikálními označeními (která se z velké části překrývají)

podobnosti, především v případě aktivního výběru podobných prvků v zobrazení, tedy tzv. *brushing*, dále v zaostření na konkrétní aspekt scény (*focusing*), and propojení objektů (*linking*) (viz tab. 2).

Interakční primitiva	
Buja et al. (1996)	(1) zaostření, (2) propojení, (3) změny náhledu
MacEachren et al. (1999)	(1) přiřazení, (2) brushing, (3) zaostření, (4) barvení mapy, (5) manipulace úhlu pohledu, (6) sekvenování
Yi et al. (2007)	(1) navigace, (2) výběr, (3) zkreslení
Keim (2002)	(1) dynamická projekce, (2) filtrování, (3) zooming, (4) zkreslení, (5) propojení a brushing
Edsall et al. (2008)	(1) zooming, (2) panning/ přacentrování, (3) změna promítnutí, (4) přesné datování, (5) zaostření, (6) změna type reprezentace, (7) změna symbolů, (8) nastavení otázky, (9) přepínání viditelnosti, (10) brushing a propojování, (11) podmiňování

Obr. 5: Primitiva založená na operátoru (dle Roth, 2012)

3.3 Primitiva na bázi operandu

Primitiva založená na operandu (*Operand-based primitives*) se nachází na pomezí Normanovy akční a evaluační části mezi stádiem provedení akce (*Executing the Action level*) a percepce stavu systému (*Perceiving the State of the System level*). Zde se jedná přímo o virtuální nebo také reálný předmět zájmu, tedy tzv. operand, reprezentovaný např. určitým typem dat, strukturou dat, časem, konkrétními objekty zájmu nebo typem ovládání. Roth tato primitiva dále dělí na typocentrická primitiva a primitiva odvislá od stavu (*type centric, state centric*, více viz Roth, 2012). Povaha těchto primitiv je různorodá a souvisí s povahou daného rozhraní (viz tab. 3).

Interakční primitiva	
Crampton (2002)	(1) data, (2) representace, (3) časová dimenze, (4) kontextualizace interakce
Keim (2002)	(1) jedno-dimenzionální, (2) dvoj-dimenzionální, (3) více dimencionální, (4) text a hypertext, (5) hierarchie a grafy, (6) algoritmy a software
Ward and Yang (2003)	(1) screen, (2) data, (3) datová struktura, (4) atribut, (5) objekt, (6) struktura vizualizace

Obr. 6: Primitiva založená na operandu (dle Roth, 2012)

Uvedené dělení bylo ověřováno také s ohledem na jeho empirické zakotvení, kde Roth (2013) prokazuje jeho aplikovatelnost v praktických úlohách. Roth jinde shrnuje (2012), že shrnutí interakčních primitiv poskytuje komplexní lexikon pro vzdělávání a praxi, tedy plní didaktické účely. Dále zdůrazňuje úlohu teoretického aparátu pro výzkumné účely, což je relevantní pro výzkum interaktivních virtuálních prostředí na mnoha úrovních. V neposlední řadě nabízená taxonomie tvoří teoretickou platformu pro design a využívání interaktivních kartografických produktů a informačních vizualizací obecně. V této linii leží problematika virtuálních geovizualizací, coby slibného nástroje pro zachycení prostorových aspektů reality ve virtuálním prostředí.

Uvedenou taxonomii je možné aplikovat ve vývoji a výzkumu výše zmíněných virtuálních 3D geovizualizací, kde uvedené dělení interakčních primitiv nabízí ucelenou typologii relevantních „tasků“ pro porovnávání např. různých typů 3D vizualizací a ovládacích zařízení pro interakci s virtuálním prostředím (viz Špriňarová, 2015; Juřík, print), popř. ověření kognitivních schopností při zpracování informačního obsahu v konkrétní geovizualizaci.

4 Závěrem

V dynamické a stále se rozvíjející oblasti virtuální vizualizace nabývá na významu ucelená taxonomie, která kategorizuje povahu interakce člověka s konkrétním rozhraním. Z tohoto úhlu pohledu je nezbytné definovat kategorie napomáhající rozdělení tzv. interakčních primitiv, aby základních stavebních jednotek interakce člověka s umělým rozhraním, a to s ohledem na možnosti ovládání daného rozhraní a jeho schopnost reprezentace žádoucí informace. Tvorba virtuálního roz-

hraní bývá určována poprvé uživatele, tedy tím, jaké operátory bude daný člověk vykonávat. Toto nastavení ovšem výrazně predeterminuje zpracování informace a tudíž je nutné je podmínit dalšímu empirickému výzkumu.

Článek implikuje poprvé po výzkumné oblasti, která by se dala označit pojmem interakční věda a reflektouje dosavadní poznatky k tomuto tématu dle Rothova dělení (Roth, 2012). V rámci článku shrnujeme taxonomii interakčních primitiv v rámci Normanova sedmi-úrovňového modelu akce a objasňujeme základní pojmy v oblasti kartografické interakce. Výčet taxonomie není v rámci tohoto článku ani zdaleka vyčerpávající a zároveň také samotné téma stále hledá své přesné ukotvení, nicméně tato exkurze tvoří základ pro pochopení základních typů interakce člověka s informačními vizualizacemi (konkrétně geovizualizacemi) ve virtuálním prostředí. Poznatky uvedené výše by měly nalézt uplatnění při tvorbě a práci s virtuálními informačními produkty.

Poděkování

Tento příspěvek vznikl v rámci projektu Vliv kartografické vizualizace na úspěšnost řešení praktických a výukových prostorových úloh (MUNI/M/0846/2015).

Literatura

- Amar, R., Eagan, J. and Stasko, J. (2005). Low-level components of analytic activity in information visualization, in *IEEE Symposium on Information Visualization*, pp. 111–117, Minneapolis, MN, Oct 23–25.
- Andrienko, G. L. and Andrienko, N. V. (1999). Interactive maps for visual data exploration, *International Journal of Geographical Information Science*, 13, pp. 355–374.
- Andrienko, N., Andrienko, G. and Gatalsky, P. (2003). Exploratory spatio-temporal visualization: an analytical review, *Journal of Visual Languages and Computing*, 14, pp. 503–541.
- Bauer, M. and Johnson-Laird, P. (1993). How diagrams can improve reasoning. *Psychological Science*, 4(6), pp. 372–378.
- Blok, C., Kobben, B., Cheng, T. and Kuterema, A. A. (1999). Visualization in relationships between spatial patterns in time by cartographic animation, *Cartography and Geographic Information Science*, 26, pp. 139–151.
- Buja, A., Cook, D. and D. F. Swayne (1996). Interactive high-dimension data visualization, *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 78–99, 1996.
- Card, S. K., Mackinlay, J. D. and Shneiderman, B. (1999). *Information Visualization: Using Vision to Think*, Morgan-Kaufmann, San Francisco, CA.
- Crampton, J. W. (2002). Interactivity types in geographic visualization, *Cartography and Geographic Information Science*, 29, pp. 85–98.
- Dykes, J. (2005). Facilitating interaction for geovisualization, in *Exploring Geovisualization*, ed. by Dykes, J., MacEachren, A. M. and Kraak, M.-J., pp. 265–291, Elsevier Science, Amsterdam.
- Edsall, R. M. (2003). Design and usability of an enhanced geographic information system for exploration of multivariate health statistics, *The Professional Geographer*, 55, pp. 146–160.
- Harrower, M. (2008). The Golden Age of Cartography is now. *Axis Maps*. <http://www.axismaps.com/blog/2008/10/the-golden-age-of-cartography-is-now/>
- Juřík, V., Herman, L., Šašinka, Č., Stachoň, Z., a J. Chmelík (Print). When the display matters: Multifaceted perspective on 3D geovisualization. *Open Geosciences*. 9(1).
- Keim, D. A. (2002). ‘Information visualization and visual data mining’, *Transactions on Visualization and Computer Graphics*, 7, pp. 100–107.
- Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11 (1), 65–100.
- MacEachren, A. M. (1995). *How Maps Work*, The Guilford Press, New York.
- MacEachren, A. M. and Monmonier, M. (1992). Geographic Visualization – Introduction, *Cartography and Geographic Information Science*, 19, pp. 197–200.
- MacEachren, A. M., Wachowicz, M., Edsall, R., Haug, D. and Masters, R. (1999). Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods, *International Journal of Geographical Information Science*, 13, pp. 311–334.
- Norman, D. A. (1988). *The Design of Everyday Things*. MIT Press
- Rierson, L. (2013). *Developing safety-critical software: a practical guide for aviation software and DO-178C compliance*. New York: CRC press.
- Roth, R. (2012). Cartographic Interaction Primitives: Framework and Synthesis. *The Cartographic Journal*, 49(4), pp. 376–395.
- Roth, R. (2013). An Empirically-Derived Taxonomy of

Interaction Primitives for Interactive Cartography and Geovisualization. *IEEE Transactions On Visualization And Computer Graphics*, 19(12), 2356-2365.

Roth, R. E. (2011). Interacting with Maps: The science and practice of cartographic interaction, PhD, Geography, The Pennsylvania State University, University Park, PA.

Špriňarová, K., Juřík, V., Šašinka, Č., Herman, L., Štěrba, Z., Stachoň, Z., Chmelík, J. a B. Kozlíková (2015). Human-computer Interaction in Real 3D and Pseudo-3D Cartographic Visualization: A Comparative Study. In Claudia Robbi Sluter, Carla Bernadete Madureira Cruz, Paulo Márcio Leal de Menezes. *Cartography - Maps Connecting the World: 27th International Cartographic Conference 2015 - ICC2015*. 1st. ed. Switzerland: Springer International Publishing, pp. 59-73, 15 s.

Wallace, T. R. (2011). A new map sign typology for the geoweb, in *International Cartographic Conference*, Paris, Jul 8.

Ward, M. and Yang, J. (2003). Interaction spaces in data and information visualization, in *Joint Eurographics/IEEE TCVG Symposium on Visualization*, pp. 1–9, Grenada, Sep 1–5.

Ware, C. (2004). *Information visualization*. San Francisco, CA: Morgan Kaufman.

Wehrend, S. (1993). Appendix B: Taxonomy of visualization goals, in *Visual Cues: Practical Data Visualization*, ed. by Keller, R. P. and Keller, M. M., p. 187, IEEE Computer Society Press, Los Alamitos, CA.

Wehrend, S., and Lewis, C. (1990). A problem-oriented classification of visualization techniques, in *1st IEEE Conference on Visualization*, pp. 139–143, San Francisco, CA, Oct 23–26.

Yi, J. S., Kang, Y. A., Stasko, J. T. and Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization, *Transactions on Visualization and Computer Graphics*, 13, pp. 1224–1231.

Zhou, M. X. and Feiner, S. K. (1998). Visual task characterization for automated visual discourse synthesis, in *Human Factors in Computing Systems Conference*, pp. 392–399, Los Angeles, CA, Apr 18–23.

Teória rečových aktov a jej implikácie

Dezider Kamhal

Centrum pre kognitívnu vedu, Fakulta matematiky, fyziky a informatiky UK,
Mlynská dolina, 842 48 Bratislava
kamhal@fmph.uniba.sk

Abstrakt

Filozofi J. L. Austin, J. R. Searle a H. P. Grice ponúkli originálne a odlišné pohľady na jazykovú prax. Ich názory ilustrujú niekoľko základných prístupov k jazykovému konaniu. Spoločným východiskom bol ich pragmatický prístup k jazyku, odlišovali sa však v dôraze, ktorý kládli na jeho konvenčnú resp. intencionálnu stránku. Všetci však upozorňovali na potrebu skúmať jazyk ako konanie a na aspekty významu/mienenia, prehliadané pri logickom alebo čisto sémantickom prístupe. Po porovnaní a zhrnutí ich názorov naznačíme možnosti ich uplatnenia v rozmanitých oblastiach komunikácie súvisiacich s modernou technológiou a pri skúmaní a modelovaní jazyka.

1 Úvod

Ústrednou ideou tvorcov a rozvíjateľov teórie rečových aktov je, že jazyk (reč) je sledom rozmanitých, viac alebo menej podobných typov konaní. Hovoriť (nie simulať hovorenie) znamená zapájať sa do aktivít riadených špecifickým druhom pravidiel (Searle, 2007). Na druhej strane, rozumieť reči znamená rozumieť konaniu podľa týchto špecifických pravidiel.

Rôzni autori zdôrazňovali rôzne zložky alebo aspekty týchto druhov konaní. Napríklad J. L. Austin vo svojich formuláciách úspešnosti performatívov kládol dôraz na konvenčnú stránku – hovoril o konvenčných postupoch, ktorých sa pridržiavajú hovorcovia a aj adresáti pri rozumení rečovým aktom. P. H. Grice pri prvej formulácii svoje psychologickej teórie významu sa zase naopak opieral o komplexnú, viaczložkovú intenciu hovorca ako základný explanačný koncept pri vysvetľovaní nielen toho, čo mieni hovorca, ale aj toho, čo konvenčne znamená nejaký výraz v jazyku (Grice, 1991).¹

V nasledujúcom texte sa obmedzím na názory a koncepcie trojice autorov, Austina, Gricea a Searlea, pretože podľa mňa typicky, každý svojským spôsobom, reprezentujú spoločnú líniu prístupu k filozofickému uvažovaniu o jazyku. Na druhej strane, ich názorové

rozdiely predstavujú krajné stanoviská vzhľadom na predpokladanú úlohu konvencí resp. intencii v rečovom konaní, prípadne pokus o syntetickejší a vyváženejší prístup. Searle zároveň ponúkol základy vedeckými metódami budovanej teórie rečových aktov.

2 Austin a „klam opisu“ ako pretrvávajúci zdroj konfúzií

J. L. Austin sa všeobecne považuje za tvorca teórie rečových aktov, aj keď nijakú teóriu v prísnom zmysle slova neponukol.² Austinove knižne vydané prednášky *How to Do Things with Words*³ začínajú zmienkou o klame, ktorému podliehali a podliehajú filozofi aj vedci pri uvažovaní o jazyku. Jeho jadrom je implicitný predpoklad, že vety jazyka, predovšetkým tie, ktoré z gramatického hľadiska patria do širokej oblasti oznamovacích viet, plnia opisnú funkciu (opisujú fakty, situácie, stav sveta a pod.) a ich kľúčovou vlastnosťou je pravdivostná hodnota. Austin nazýva vyslovenia viet, ktoré plnia túto funkciu a ktoré hodnotíme z hľadiska pravdivosti či nepravdivosti „konštatívmi“, a ďalej sa sústredí na výpovede, ktoré plnia odlišné funkcie. Tie explicitne charakterizuje ako konanie resp. vykonávanie činov a nazýva ich „performatív“. Na rozdiel od konštatívov ich môžeme posudzovať z hľadiska úspešnosti či neúspešnosti. Práve Austinom sformulované podmienky úspešnosti performatívov prezárdzajú hranice jeho uvažovania. Vo všetkých sa predpokladá, že pri hovoreni sa opierame o konvenčné postupy. Akékoľvek porušenie týchto postupov viedie k nevydarenému performatívu, ba v prípade prvých štyroch z jeho šiestich podmienok k ich nulitnosti, úplnému nezdaru.

Predpoklad konvenčnosti nášho jazykového konania je sám osobe v poriadku. Vyžaduje si však doplnenie, pretože čisto konvencionalistický pohľad nedokáže vysvetliť dve samozrejmosti: existenciu nekonvenčného (a úspešného!) jazykového konania a vznik

² Uvedomoval si rozmanitosť a neostrosť hraníc konvencí, ktorých sa pridržíame pri jazykovom konaní. Navrhol však terminológiu pri skúmaní aspektov rečových aktov, ktorá sa ujala.

³ Slovenský preklad: Austin, J. L. (2004): *Ako niečo robit slovami*, Kalligram

¹ Ide o jeho dôležité rozlíšenie medzi utterer's meaning a expression meaning.

samotných konvencií. Aj sama konvenčnosť jazykového konania však predstavuje výzvu pre odborníkov z rôznych oblastí zaobrájúcich sa jazykom. Aká je povaha týchto konvencií resp. pravidiel?⁴ Možno ich zachytiť vo forme striktných definícii?

Austin v priebehu spomenutých prednášok čoskoro prezradil pointu svojho prístupu. Spočívala v tom, že všetky druhy výpovedí, teda aj konštatóvy, sú vlastne rečovými aktami, ktoré – ako napokon každé konanie – môžu byť aj nevydarené. Jednotkou skúmania jazyka je tak „celkový rečový akt v celkovej rečovej situácii“ (Austin, 2004, 57). Táto formulácia naznačuje obrovský priestor faktorov, ktoré ovplyvňujú úspešnosť nášho hovorenia a ktoré pri jeho analýze treba bráť do úvahy.

Ked' si Austin nakoniec opäť položil otázku, v koľkých rôznych zmysloch môžeme povedať, že niečo robíme, ked' hovoríme, rozlíšil viacero aspektov rečových aktov: lokučný, ilokučný a perlukučný. Každý rečový akt možno skúmať a opisovať z týchto hľadísk. Do lokučného aspektu zahrnul aj to, čo by sme spojili so syntaktickou a sémantickou stránkou výrazu a jeho použitia. Za najdôležitejšie pritom možno povaľať to, že upozornil na ilokučný aspekt/akt ako neodmysliteľnú významovú zložku rečového konania. Ak človek nepochopí tzv. ilokučnú účinnosť konkrétneho rečového aktu, neporozumie mu.

Austin sa venoval aj problematike možných kritérií performatívov. V súvislosti s tým uvažoval o výskytu tzv. performatívnych slovies v explicitných performatívoch a ich opisoch a o ďalších gramatických kritériách. Identifikácia performatívnych slovies pri prehľadávaní korpusových textov sa (aj napriek ich nespoľahlivosti) využíva ako jeden z účinných identifikačných prostriedkov ilokúcií pri prehľadaní textových databáz.

Austin opakovane pripomínal „neostrost“ konvencií, ktorími sa – väčšinou implicitne – riadime pri vyslovovaní výpovedí, pri pokusoch vysloviť ich, a pri filozoficky motivovanej pragmatickej analýze niektorých druhov výpovedí vždy poukazoval na konkrétnu situácie a okolnosti rečového konania. Neostrosť konvencií, presnejšie konvenčných procedúr, ako ich neoddeliteľná charakteristika znemožňuje sformulovať akési analytické definície, ktoré by jednoznačne určovali význam výrazov, ktorími opisujeme jednotlivé druhy rečového konania.

3 Griceova psychologická teória významu

Namiesto abstraktných otázok týkajúcich sa významu nejakého druhu výrazov alebo výpovedí si P. H. Grice položil oveľa „prízemnejšiu“ otázku: Čo to znamená, že niekoľko niečo mieni vyslovením nejakého výrazu? V relatívne krátkom článku „Meaning“⁵ tak predstavil

⁴ Searle rozlišuje medzi konvenciami a pravidlami.

⁵ V slovenskom preklade správne preloženom ako „Mienenie a význam“.

ambiciozny program, ktorý mal nakoniec vyústiť aj do odpovedí na abstraktné otázky týkajúce sa (konvenčného) významu jazykových výrazov. Predpokladal, že vysvetlenie toho, čo kto mieni vyslovením výrazu, bude východiskom vysvetlenia toho, čo znamená, že nejaký výraz má význam v nejakom jazyku.

Jeho pôvodná odpoveď na otázku, čo znamená, že hovorca niečo mieni vyslovením nejakého výrazu znala nasledovne: „„A niečo mienil prostredníctvom *x*“ je (priľahlé) ekvivalentné „Zámerom *A* bolo vyvolat’ vyslovením *x* určitý účinok v poslucháčoch rozpoznaním tohto zámeru“ (Grice, 1992, 176). Dôležité je upozorniť, že jeho vysvetlenie *mienenia* sa netýkalo len tvrdení, ale aj iných druhov výpovedí. Mienenie teda vysvetľoval zámerom (intenciou) hovorcu, ktorá – ako si sám uvedomoval – má tri zložky. Po prvej, hovorca má

v úmysle dosiahnuť nejaký účinok na adresáta, po druhé, chce, aby adresát tento úmysel rozpoznal a po tretie, chce zamýšľaný účinok dosiahnuť rozpoznaním jeho úmyslu. Nech sa už následne jednotlivé zložky griceovskej zloženej intencie akokoľvek kritizovali a nech Grice svoje vysvetlenie akokoľvek modifikoval, podstatné bolo uvedomenie si faktu, že pre pochopenie jazykového konania hovorcu (jeho výpovedného aktu) je podstatné identifikovať intencionálnu zložku tohto konania.

Rozlúšlenie medzi tým, čo hovorca mieni vyslovením výrazu a tým, čo výraz (konvenčne) v jazyku znamená plodne využíva aj pri analýze častých prípadov tzv. implikatív.⁶ Predložil logickú rekonštrukciu postupu, akým sa adresát „dovtípi“, čo v takýchto prípadoch hovorca mienil vyslovením nejakého výrazu. V jeho vysvetlení hral klúčovú úlohu princíp spolupráce a s ním súvisiace maximy. Inšpiroval tým vznik ďalších teórií komunikácie pracujúcich s alternatívnymi ústrednými princípmi (napr. princípom zdvorilosti, princípom relevantnosti).

4 Searlova syntetizujúca teória

Searle svoju dnes už klasickú prácu *Rečové aktu* charakterizoval ako postupné overovanie jej hlavnej hypotezy – že hovorenie je konanie podľa špecifického druhu pravidiel.

Jeho teóriu rečových aktov možno chápať ako pokus prevziať pozitíva konvencionalistického aj intencionalistického vysvetlenia jazykového konania a na ich základe vybudovať plnohodnotnú formalizovateľnú teóriu. Searle kriticky rozvíja najmä Austinove a Griceove idey. Východiskom jeho vysvetlenia významu výrazu/mienenia hovorcu je Griceova formulácia, ktorú opravuje a dopĺňa. Správne kritizuje Gricea za to, že v jeho vysvetlení intencie hovorcu sa hovorí o perlukučnom účinku v austinovskom zmysle. Prevedčivo ukazuje, že hovorcovi ide namiesto toho pre-

⁶ Termín zaviedol Grice. Ide o taký jazykový jav, keď, voľne povedané, hovorca mieni niečo iné ako povie.

dovšetkým o to, aby adresát pochopil, akú účinnosť má mať jeho výpoved, teda o ilokučný účinok. Navýše, Griceovu formuláciu dopĺňa o konvenčnú zložku – hovorca chce dosiahnuť zamýšľaný účinok (spravidla) prostredkami, ktoré sa na dosahovanie príslušného účinku *konvenčne* používajú.

V súvislosti s druhou opravou predložil myšlienkový experiment, ktorý mal ilustrovať ideu, že – voľne povedané – hovorca nemôže na dosiahnutie zamýšľaného účinku na adresáta povedať hocičo. Je však zaujímavé, že aj keď v myšlienkovom eksperimente opísal situáciu tak, aby vyznelo absurdne povedať, že vyslovením náhodnej nemeckej vety (verša z básne) chcel zajatý americký dôstojník presvedčiť svojich talianskych väzniteľov, že je nemecký dôstojník, za istých okolností to nie je nemysliteľné. To späť vyzdvihuje pôvodné Griceovo čisto intencionalistické vysvetlenie. Treba však uznať, že Searlovo syntetizujúce vysvetlenie sa týka drívnej väčšiny našich výpovedí. Ak v komunikácii nesledujeme iné ciele, na dosahovanie svojich úmyslov čo najviac využívame konvenčné prostriedky.

Na druhej strane, od Austina prebral pojmom ilokučného a perlukučného aktu a ideu, že hovorenie je konaním podľa pravidiel. Na rozdiel od Austina však nielenže videl zmysel v budovaní teórie ilokučných aktov, ale pokladal to za jednu z dôležitých úloh filozofie. V tomto ohľade rozhodne nebol Austinovým pokračovateľom.

Možno to vyznie paradoxne, ale problematicosť Searlových postupov sa dá vidieť najmä v jeho konkrétnych pokusoch ponúknutí základné pojmy, ktoré by umožnili vytvorenie formálnej teórie rečových (ilokučných) aktov. Spolu s D. Vandervekenom dokonca ponúkol základy tzv. ilokučnej logiky, ktorá je podľa nich „logickou teóriou ilokučných aktov“ a jej hlavným cieľom je „formalizovať logické vlastnosti ilokučných účinností“ (Searle A Vanderveken, 1985, 1).

Pri vytváraní teórie rečových aktov sa otvorené opieral o vedecké metódy (abstrakcia, idealizácia, generalizácia), čo mu umožnilo položiť základy symbolického vyjadrenia štruktúry ilokučných aktov. Jeho jednoduchý návrh na všeobecnú formu ilokučných aktov $F(p)$, kde za premenná F dosadzujeme prostriedky indikujúce ilokučnú účinnosť a hodnotami premennej p sú výrazy pre propozície poukazujúce na slabiny formalizácie pri skúmaní jazyka. Táto ním navrhnutá jednoduchá forma sa totiž nehodí na mnohé druhy a prípady ilokučných aktov a nastoľuje otázku, aká vysoká je cena, ktorú treba za pokusy o formalizáciu uvažovania o skutočných jazykoch zaplatiť.

V zmazávaní hraníc medzi odlišnými rečovými aktmi pre potreby formalizácie Searle zašiel naozaj ďaleko. Kým povedať, že propozičný obsah vyjadrený vetou „Je neskoro“ je rovnaký vo všetkých výpovediach s rôznou ilokučnou účinnosťou, v ktorých túto vetu vyslovujeme, znie ešte vcelku prirodzene, menej prirodzene už znie, že výpovede „Samo prestal

fajčiť“, „Samo, prestaň fajčiť!“, „Prestal Samo fajčiť?“ či „Pán Samuel Martin prestal fajčiť“ majú rovnaký propozičný obsah. (Prípadne, že pri ich vyslovení hovorca vykoná rovnaký propozičný akt (Searle, 2007, 40–41). O to viac, keď sa dodá, že pri každom vyslovení tejto údajnej spoločnej propozície v rôznych druhoch rečových aktov (resp. pri každom propozičnom akte) hovorca predkuje vlastnosť „fajčenia“ tej istej osobe s menom Samo. Prípadne, že „...o človeku, ktorý tvrdí, že Sokrates je múdry, o človeku, ktorý sa pyta, či je Sokrates múdry a o človeku, ktorý Sokrata žiada, aby bol múdry, možno povedať, že vznáša otázku, či je Sokrates múdry (či výraz „múdry“ je – alebo v prípade žiadosti bude – o Sokratovi pravdivý“ (Searle, 2007, 172).

Dá sa povedať, že Searle zaplatil privysokú cenu za snahu zachytiť formálnu štruktúru rečových aktov – natoľko sa vzdialil od významu takých výrazov, ako je pomerne bežný výraz „pripisovať (niekomu, niečomu) vlastnosť“, „opisovať (niečo, niekoho)“, „vzniest otázkou“ a pod., že pri charakterizácii niektorých druhov rečových aktov, ktoré sa opierali o takto predefinované výrazy sa vzdialil aj od ich bežného chápania. Prípady úspešného rečového konania, ktoré by kompetentní hovorcovia bez problémov opísali pomocou ilokučných slovies (napr. „sľubil“, „poprosil“, „pričkal“, „ospravedlnil sa“ atď.) sa tak dostali mimo jeho teórií.⁷ Nedostatky jeho teórie priponínajú, akým komplexným fenoménom je naše jazykové konanie a naznačujú neprekročiteľné hranice pri pokusoch opísť ho pomocou formalizovaného jazyka.⁸

Zdá sa, že príliš ďaleko zašiel aj v zmazávaní hraníc medzi výpovedami patriacimi do jedného druhu rečových aktov. Všimnime si jeho vzorovú analýzu sľubovania ako druhu rečového aktu a jeho vlastné poznámky k metódam, ktoré spája so svojou analýzou! Searle napríklad zámerne vyniechal zo svojich analýz neexplicitné sľuby, teda také, ktoré neobsahujú príslušný tvar ilokučného slovesa „sľubovať“. Obmedzujúce a nedostatočne zdôvodnené sú aj ďalšie jeho predpoklady, ktoré sa v nejakej podobe objavia vo formuláciách nevyhnutných a postačujúcich podmienok sľubovania. Pripomeňme si, kolko rozmanitých rečových aktov možno vykonať obyčajným prikývnutím alebo pritakaním. Sú to okrajové alebo druhoradé prípady rečového konania?

Kým Austin až prehnane zdôrazňoval úlohu jazykových konvencii pri rečovom konaní a len okrajovo sa zmienil o úlohe hovorcových intencií pri niektorých druhoch performatívov, Searle vychádzajúci z Griceovej psychologickej teórii významu považoval intenciu hovorcu za neodmysliteľnú súčasť všetkých výpovedí. Umožnilo mu to odstrániť hlavné slabiny Austinovho

⁷ Na neadekvátnosť Searlových nevyhnutných a postačujúcich podmienok pre sľubovanie ako druh ilokučného aktu poukázal Alston (2000).

⁸ Pod neprekročenými hranicami mám na mysli to, že akákoľvek teória jazyka je sama *iným* skonštruovaným jazykom.

konvencionalistického chápania rečového konania – neschopnosť vysvetliť, ako môže adresát porozumieť rečovému konaniu, ktoré sa neopiera o konvencie, čiže aj neschopnosť vysvetliť samotný vznik jazykových konvencí.

5 Niektoré implikácie teórie rečových aktov

Zhrňme si niektoré viac-menej samozrejmé fakty o rečovom konaní, na ktoré priamo alebo nepriamo upozornili Austin, Grice a Searle. V prvom rade, vyslovenie/napísanie slov alebo zložených výrazov v reálnych prípadoch komunikácie je vždy konaním, v prípade úspechu vykonaním komplexného činu, ktorý sice možno skúmať z rôznych aspektov, abstrahujúc od jeho rôznych vlastností a okolností (napr. z tzv. čisto sémantického hľadiska odhliadajúceho od situačného kontextu), ale netreba zabúdať na jeho komplexnosť.

Teória rečových aktov ďalej poukazuje na nasledovné stránky rečového konania, na ktoré sa predtým upozorňovalo len sporadicky alebo sa úplne prehliadali:

- kľúčovú úlohu konvencií, konvenčných procedúr, ktorými sa riadia a na ktoré sa implicitne odvolávajú používatelia jazyka
- primeranosť hovorcov príslušnej konvenčnej procedúre, ich vlastností v širokom zmysle slova alebo ich sociálnej roly, inštitucionálneho postavenia a pod. ako dôležitú podmienku úspešnosti ich rečového konania⁹
- dôležitú úlohu hovorcovho zámeru pri niektorých druhoch výpovedí¹⁰ – čiže intencie hovorca spojenej s vyslovením nejakého výrazu, predpokladanej a adresátkami očakávanej na základe konvencii; pričom hovorca nemusí mať konvenčne očakávané a požadované zámery a môže tak konvencie zneužívať (napr. klamať, falošne sľubovať a pod.)
- fakt, že rečové konanie je takto spojené zo záväzkami, ktoré na seba hovorca preberá, pričom tieto záväzky určujú aj primeranost či správnosť jeho následného rečového alebo iného konania¹¹
- jednotkou skúmania jazyka je rečový (ilokučný) akt, prípadne, ako to zdôrazňoval Austin, „celkový rečový akt v celkovej situácii“ (Austin, 2004, 57).

⁹ Napríklad, hocikto nemôže hocikomu rozkazovať. Pre úspešnosť rozkazu ako druhu rečového konania je teda podstatné, aby rozkazujúca osoba bola formálnou alebo neformálnou autoritou príslušného druhu.

¹⁰ Podľa Searla, Gricea a ďalších pri všetkých typoch výpovedí.

¹¹ Možno ich vyjadriť napríklad výpovedami typu „Sľuby sa majú plniť“, „Nemali by sme si protirečiť“ atď.

– pokus vykonať rečový akt nemusí byť vždy úspešný a jeho úspešnosť či neúspešnosť súvisí s podmienkami pre jednotlivé druhy rečových aktov

– rečovými aktami sú aj výroky (tvrdenia, konštatívy) a výpovede tohto druhu takisto môžu byť ako činy nevydarené, a to viacerými zaujímavými spôsobmi

– dôležitým faktom, meniacim pohľad na predmet sémantiky je, že tzv. ilokučná účinnosť výpovede je neoddeliteľnou zložkou jej významu – bez pochopenia toho, čo zamýšľal hovorca urobiť pri výpovedi (nie jej prostredníctvom)¹² nemožno hovoriť o tom, že adresát mu porozumel; nestaciť priradiť vete (alebo slovu) propozičný obsah (alebo doslovny význam) alebo pravdivostné podmienky, pretože tá istá abstraktne chápaná veta môže byť použitá na rôzne druhy výpovedí¹³ a v tom zmysle mať odlišné významy

– komunikácia obsahuje indikátory ilokučnej účinnosti, ktorých analýza môže byť užitočná aj v oblasti informatiky a komunikácie využívajúcej nové technológie

– adekvátny pohľad na jazyk nepredpokladá automaticky, že hlavným typom výpovedí sú konštatívy a že prvoradou funkciou jazyka je opisovať fakty

6 Teória rečových aktov a jej aktuálne a možné uplatnenie

Uvediem predovšetkým niektoré dôsledky relevantné pre „komunikáciu“ s počítačom alebo pre počítačom sprostredkovanú komunikáciu. Komunikácia chápaná prirodzene ako sled rečových aktov (v ktorých nemusí ísť o „výmenu informácií“) vykonávaných ľuďmi s ich poznávacími schopnosťami a na pozadí spoločnej aj individuálnej zásoby rozmanitých poznatkov určuje hranice, v akých možno rozumne hovoriť o komunikácii človeka s počítačom. Je otázne, či už samo slovo „komunikácia“ nie je v tejto súvislosti zavádzajúce. Počítač tu nikdy nebude partnerom v komunikácii alebo, povedané miernejšie, rovnocenným partnerom, pokiaľ nebude možné hovoriť o jeho poznávaní a poznatkoch v aspoň približne (a aspoň v niektorých aspektoch) podobnom zmysle slova ako to robíme u ľudí. Je dôležité rozlišovať medzi metaforickým hovorením o konaní, poznatkoch a poznávaní v prípade počítačov a hovorením o konaní, poznatkoch a poznávaní v prípade ľudí.

¹² V bežnom jazyku u Austina „in saying“, nie „by saying“.

¹³ Medzi často uvádzané príklady viet, ktorých vyslovenie v rôznych rečových situáciach môže byť vykonaním rozdielnych rečových aktov patria napr. vety „Dvere sú otvorené“, „Je neskoro“, „Som hladný“. Austin (2004) uvádzá ako príklad vetu „Zastrel ju!“

Je zrejmé, že idey teórie rečových aktov (TRA) majú a môžu mať najrozsiahlejšie uplatnenie v oblastiach priamo sa týkajúcich nejakej formy komunikácie. Predovšetkým vo sfére komunikácie/interakcie človeka s počítačom a počítačmi sprostredkovanej komunikácie. Prehľadný opis využitia TRA v oblasti interakcie človeka s počítačom možno nájsť v Dix, A. a kol. (2004), najmä v kapitole „Komunikácia a modely spolupráce“.

Kedže pole aktuálneho a možného uplatnenia TRA je veľmi široké, uvediem len niektoré príklady. Priznávam, že ma prekvapil rozsah, v akom sa koncepty a idey rôznych variantov TRA či filozofických teórií komunikácie pri tvorbe a vývoji aplikácií a softvéru všeobecne, používateľského rozhrania operačných systémov, internetových stránok, v oblasti informatiky, kognitívnej vedy atď. už používajú.

Ďalšou oblasťou je dizajnovanie používateľský prístupných informačných systémov, prípadne zjednocovanie ich vzájomnej komunikácie. Tu už môže ísť o komunikáciu, ktorej sa človek priamo nezúčastňuje. Napríklad v Mihálik, I., Rimassa, G. (2007) sa uvádzá: „Sémantická komunikácia definuje štyri performatív, ktoré pokrývajú najdôležitejšie prípady pri komunikácii: REQUEST na vyjadrenie požiadavky na akciu, NOTIFY na vyjadrenie prenosu informácie o nastanej udalosti, ASK-IF na vyjadrenie uzavretých (áno/nie) otázok a ASK-WHICH na vyjadrenie otvorených otázok. Tieto štyri performatív boli zvolené ako najzákladnejšie. S ohľadom na FIPA sémantická komunikácia obsahuje dve základné performatív a dve najbežnejšie z množiny odvodených“ (Mihálik a Rimassa, 2007).

Významným príkladom využitia teórie rečových aktov sú aktivity spomenutej organizácie FIPA.¹⁴ V Programovom vyhlásení FIPA (The Foundation for Intelligent Physical Agents [FIPA], 2003) sa doslova píše: „Základnou ideou FIPA je, že prostredníctvom kombinácie rečových aktov, predikátovej logiky a verejných ontológií môžeme poskytnúť štandardné spôsoby interpretovania komunikácie medzi agentami spôsobom, ktorý rešpektuje *zamýšľany*¹⁵ význam komunikácie.“ Vo FIPA Communicative Act Library Specification (FIPA, 2003) sa definuje 22 komunikačných aktov explicitne inšpirovaných aj Searlovou prácou *Rečové akty*.

Môžeme sa stretnúť aj s využitím TRA v tzv. korpusovej lingvistike, ktorá využíva rýchle spracovávanie rozsiahlych, často špecificky zameraných súborov textov.¹⁶ Pre zaujímavosť, časť publikácie Taavitsainen, I., Jucker, A. H., Tuominen, J. (2014) s príznačným

¹⁴ FIPA (The Foundation for Intelligent Physical Agents) – medzinárodná organizácia, ktorá sa venuje podporovaniu odvetvia inteligentných agentov verejne prístupným rozvíjaním špecifikácií podporujúcich interoperabilitu medzi agentmi a na agentoch založenými aplikáciami.

¹⁵ Kurzíva D. K.

¹⁶ Existujú napr. korpusy hovorového jazyka, súkromnej korepondencie, poézie a pod.

názvom „Výpovede a dialógy“ obsahuje kapitoly o skúmaní verbálnej agresivity v anglických historickej textoch a o robení komplimentov či poklón v americkej angličtine, čo si v oboch prípadoch vyžaduje identifikovanie zamýšľaného ilokučného účinku. Na problémy, na ktoré naráža skúmanie a identifikovanie rečových aktov nielen v korpusovej lingvistike, poukazuje nasledujúci citát: „Menej konvenčné rečové akty¹⁷ napríklad žiadostí, komplimenty či urážky, sa však môžu vyskytovať v nekonečnom počte rôznych realizácií a je oveľa ľahšie hľadať ich, ak už úplne nevzdujú automatickým vyhľadávacím technikám“ (Taavitsainen a Jucker, 2007, 107).

Ďalšími zaujímavými príkladmi sú analýza e-mailovej korepondencie z rôznych aspektov¹⁸, online alebo elektronickej komunikácie rôznych druhov (napr. IRC komunikácie (Herring, 1996) a emotikonov a ich funkcií (Dresner a Herring, 2014)¹⁹.

7 Záver

V tomto príspevku som sa nijako nepokúsil odpovedať na otázky, ktoré sú podľa mňa klúčové pre pochopenie toho, v čom môžu byť filozofické analýzy jazyka užitočné pre jednotlivé vedné odbory: Kde sa začína a kde končí analógia medzi ľudskou komunikáciou a komunikáciou súvisiacou s modernými technológiami? Podobne, analógia medzi konštruovaním modelov vo vede a filozofickou analýzou jazyka? V akom zmysle možno rozumne hovoriť o rečovom konaní v prípade počítačov? Odpovede na tieto otázky sú dôležité aj pre pochopenie cieľov, možností a hraníc uplatnenia rôznych variantov teórie rečových aktov v informatike všade tam, kde je reč o jazyku a komunikácii v najširšom zmysle.

Pod'akovanie

Tento príspevok vznikol s podporou grantovej agentúry KEGA v rámci projektu č. 017UK-4/2016.

Literatúra

- Alston, W. P. (2000). *Illocutionary Acts and Sentence Meaning*, Cornell University Press.
- Austin, J. L. (2004). *Ako niečo robiť slovami*, Kalligram.
- Dix, A. a kol. (2004). *Human-Computer Interaction*, 3. vyd., Pearson Education Limited.

¹⁷ V porovnaní napr. s pozdravmi.

¹⁸ Etické aspekty si všíma Rooksby (2002), pričom vychádza z Austinovej koncepcie.

¹⁹ Autori ponúkajú výklad, podľa ktorého emotikony nie sú len indikátormi emócií, ale aj indikátormi ilokučnej účinnosti.

Dresner, E., Herring, S. C. (2014). Emoticons and Illocutionary Force, in Riesenfeld, D., Scarafale, G. (eds), *Perspectives on Theory of Controversies and the Ethics of Communication*, Springer.

Grice, H. P. (1975). Logic and conversation, *Studies in the way of words*, Harvard University Press, 22 – 40.

Grice, P. H. (1991). Meaning, *Studies in the way of words*, Harvard University Press, 213 – 223.

Grice, P. H. (1992). Mienenie a význam, *Filozofia prirodzeného jazyka*, Bratislava, Archa.

The Foundation for Intelligent Physical Agents. (2003). FIPA's Mission Statement. Retrieved March 31, 2017, from <http://www.fipa.org/about/mission.html>

Herring, S. C. (1996). Two Variants of an Electronic Message Schema, Herring, S. C (ed.), *Computer-mediated communication: linguistic, social, and cross-cultural perspectives*, John Benjamins Publishing Company, 81 – 106.

Mihálik, I., Rimassa, G. (2007). Sémantická komunikácia v informačných systémoch, *Informačné technológie a knižnice*, 1/2007 [online]. [cit: 2017-03-31] Dostupné na <http://itlib.cvtisr.sk/>

Rooksby, E. (2002). *E-mail and Ethics. Style and ethical relations in computer-mediated communication*, Routledge.

Searle, J. R. (2007). *Rečové akty*, Kalligram.

Searle, J. R., Vanderveken, D. (1985). *Foundations of Illocutionary Logic*, Cambridge University Press.

Taavitsainen, I., Jucker, A. H. (2007). Speech act verbs and speech acts in the history of English, Taavitsainen, I., Fitzmaurice, S. M. (eds.) *Methods in historical pragmatics*, Mouton de Gruyter.

Taavitsainen, I., Jucker, A. H., Tuominen, J. (2014). *Diachronic Corpus Pragmatics*, John Benjamins Publishing Company.

Vplyv iracionálneho presvedčenia na úspešnosť v riešení úloh u finančných profesionálov

Lenka Kostovičová, Katarína Dudeková, Matúš Konečný

Ústav experimentálnej psychológie CSPV SAV

Dúbravská cesta 9, 84101 Bratislava

lenka.kostovicova@savba.sk, katarina.dudekova@savba.sk, matus.konecny@savba.sk

Abstrakt

Podľa nedávnej štúdie iracionálne presvedčenie o šťastí zlepšuje výkon v motorických a kognitívnych úlohách, no replikácie zistenia spochybnila. My sme zrealizovali experiment s 52 finančnými profesionálmi. Súčasťou inštrukcie pre experimentálnu skupinu bolo navodenie iracionálneho presvedčenia o šťastí. Participanti pred vyriešením 7 úloh odhadli svoju úspešnosť a na záver vyjadrili, či veria v šťastné predmety a symboly. Experimentálna skupina nepodala lepší výkon v žiadnej úlohe, ani celkovo. Participanti bez apriórneho iracionálneho presvedčenia však boli výrazne úspešnejší vo väčšine úloh, a dosiahli významne lepšie skóre. Zdôrazňujeme potrebu rozvoja kritického myslenia a racionality, s cieľom podpory optimálnych osobných a profesionálnych rozhodnutí.

1 Iracionálne presvedčenia

Literatúra o psychológiu posudzovania a rozhodovania je plná príkladov iracionálnych úsudkov a volieb (e.g., Ariely, 2009; Kahneman, 2011; Stanovich, 2009, 2011). Presvedčenia sú súčasťou kognícii jednotlivca, pričom miera, do akej korešpondujú s reálnym svetom predstavuje úroveň jeho epistemickej racionality (Stanovich, 2012). Iracionálne presvedčenia, ako napríklad viera v paranormálne javy, astrológiu či homeopatiu, sú súčasťou tzv. kontaminovaného mindware-u. Ide o obsah ľudskej myслe, ktorý vedie k maladaptívnomu správaniu (Čavojová a spol., 2016; Stanovich, 2009).

Súčasťou iracionálnych presvedčení sú aj povery o šťastí. Najrozšírenejšie lokálne príklady súvisia s čierrou mačkou, štvorlístkom, komínárom či šťastnými amuletmi. Viera v šťastné kusy oblečenia je napríklad pomerne rozšírená medzi športovcami (Damisch a spol., 2010), preto je fenomén niekedy nazývaný ako "efekt šťastných ponožiek". Ak veríme, že nám čokoľvek šťastie prináša a iné nás oň, naopak, Oberá, no logika a empirická evidencia tento predpoklad nepodporujú, ide o iluzórnu koreláciu. Teda, vnímame vzťah tam, kde nie je - napríklad sa domnievame, že vďaka "šťastnému" peru lepšie dopadneme v teste.

1.1 Východiskové štúdie

V roku 2010 publikovali Damisch a spol. výskum s prekvapivými výsledkami. Vo viacerých blogoch a v neskoršej replikačnej štúdií (Calin-Jageman a Caldwell, 2014) ich označovali za kontra-intuitívne a zarážajúce. Autori totiž zistili, že navodenie iracionálneho presvedčenia o šťastí zlepšilo výkon participantov v motorických i kognitívnych úlohách. V experimentoch s kognitívnymi úlohami identifikovali i mediátor efektu - vyššiu seba-účinnosť (tzv. "self-efficacy", t.j. viera vo vlastné schopnosti), ktorá vedie k stanoveniu si náročnejších cieľov a väčšej vytrvalosti pri ich plnení.

Štúdia Damisch a spol. (2010) pozostávala zo štyroch experimentov. Prvé dva sa týkali motorických zručností, tretí pamäte a štvrtý úlohy s anagramami (tvorba slov). V prvom navodili presvedčenie vetou: "Toto je vaša loptička. Dosiaľ sa ukázala byť šťastnou loptičkou" a v druhom povzbudením: "Držím vám palce". Posledné dva experimenty boli založené na princípe, že si všetci doniesli svoj šťastný predmet (o ktorom písali v rámci zdanivo nesúvisiacej časti výskumu) a situácia bola zinscenovaná tak, že len členovia experimentálnej skupiny mali šťastný predmet pri sebe počas hlavnej časti experimentu (Damisch a spol., 2010). Zistenia sú o to nečakanejšie, že experimenty boli z hľadiska veľkosti vzorky poddelené ($n = 28-51$).

Calin-Jageman a Caldwell (2014) prvý z experimentov dvakrát zreplikovali s dostatočne veľkou vzorkou na základe výpočtu sily (*power calculation*). Ich výsledky však vplyv iracionálneho presvedčenia nepodporili. Do meta-analýzy zahrnuli okrem štyroch východiskových experimentov i dva staršie experimenty pôvodného autora (Damisch, 2008), ďalej dva vlastné experimenty, jeden experiment kolektívú Lee a spol. (2011; rovnaká motorická úloha) a dva experimenty kolektívú Aruguete a spol. (2012; úlohy na uvažovanie). Meta-analýza vyhodnotila veľkosť efektu ako $d = 0,4$, no vysoká heterogénnosť experimentov obmedzuje možnosť spoľahlivej interpretácie.

1.2 Výskumný cieľ a otázky

V našej práci sme sa zamerali na kognitívne úlohy, v ktorých ľudia často zlyhávajú. V pilotnej štúdie Nedorostová (2017) použila indukciu, založenú na šťastnom symbolu. Manipulácia zvýšila vnímanú sebaúčinnosť, no mala odlišné efekty na mužov a ženy. Výkon mužov v dôsledku intervencie významne klesol, zatiaľ čo už žien zásadne stúpol. My sme sa rozhodli experimentálnu manipuláciu založiť na rovnakom princípe, s využitím iného šťastného symbolu, ktorý bol pre účel online zberu dát vhodnejší. Keďže výsledky pilotnej štúdie boli poznáčené efektom podlahy, kedy mali niektoré kognitívne úlohy nulovú úspešnosť, oslovoili sme skupinu finančných profesionálov. Domnievali sme sa, že vďaka práci v oblasti financií budú numerické problémy pre nich menej náročné. Uspôsobili sme ich tak, aby boli bližšie ich profesií, a teda viac ekologicky valídne.

Vybrali sme problémy, ktoré reprezentujú kľúčové kompetencie v oblasti financií. Prvá úloha (upravená z Toplak a spol., 2014) sa týkala kognitívnej reflexie, schopnosti zapojiť analytické myslenie a potlačiť automatické intuitívne odpovede (Frederick, 2005). Druhá úloha zistovala sklon k jednej z najrobustnejších kognitívnych odchýlok - k omylu konjunkcie, kedy človek označí pravdepodobnosť dvoch javov ako vyššiu než pravdepodobnosť jedného z nich (Tversky a Kahneman, 1983). Ďalšie dva problémy overovali finančnú (Bačová a spol., 2017) a numerickú gramotnosť (Lipkus a spol., 2001). Posledné tri úlohy sa týkali výpočtu pravdepodobnosti - kumulatívnej a podmienejnej (upravené z McCloy a Beaman, 2005; Bertrand in Falk a Bar-Hillel, 1982; posledná úloha vymyslená autormi príspevku).

Narozdiel od experimentov 3 a 4 autorov Damisch a spol. (2010) sme chceli mať v súbore zastúpených ľudí s rôznym predchádzajúcim presvedčením o funkčnosti šťastných symbolov a predmetov. Zaujímalo nás predovšetkým, či navodenie presvedčenia o šťastí ovplyvní výkon finančných profesionálov v rámci numerických úloh. Zároveň sme chceli overiť rolu apriórnych presvedčení: bude úspešnosť v riešení kognitívnych problémov rovnaká u ľudí, ktorí v šťastné symboly veria a u tých, ktorí v ne neveria?

2 Metódy

2.1 Participanti a dizajn

Zo skupiny zúčastnených sme vylúčili 4 respondentov s profesiou mimo finančnej domény. Následne sme analyzovali dáta od 52 finančných profesionálov vo veku 22 až 54 rokov ($M = 34,1$; $SD = 7,3$). Muži tvorili 65,4% ($n = 34$) a ženy 34,6% ($n = 18$) súboru. Profesie participantov boli rôznorodé, medzi inými: účtovník, ekonóm, finančný analytik, finančný agent, privátny finančný konzultant a investičný poradca. Participanti

boli náhodne rozdelení do kontrolnej ($n = 26$) a experimentálnej skupiny ($n = 26$). Experimentálna manipulácia spočívala v navodení iracionálneho presvedčenia o šťastnom číslе sedem (viď ďalší odsek).

2.2 Materiály a procedúra

Inštrukcia pre experimentálnu skupinu znala nasledovne: "O chvíľu budete riešiť 7 úloh rôznej náročnosti, týkajúcich sa finančnej domény. Veľa ľudí verí, že im číslo 7 zabezpečí šťastie a úspech vo všetkom, do čoho sa pustia. Odhadnite, koľko zo siedmich úloh vyriešite správne (0 až 7): ____". V prípade kontrolnej skupiny bola druhá veta vynechaná. Participantov sme taktiež informovali, že úlohy boli pre potreby výskumu značne zjednodušené.

Po odhadnutí svojej úspešnosti participanti riešili úlohy, zoradené podľa predpokladanej stúpajúcej náročnosti. Ďalej uvádzame znenie úloh so správnymi odpoveďami:

1) **Kognitívna reflexia [KR]:** "Súrodenci Ján a Mária sporia na bicykel. Ján by naň nasporil za 60 dní. Mária by naň nasporila za 120 dní. Ako dlho by im trvalo nasporiť na bicykel, keby sa rozhodli kúpiť si spoločný? ____ dní" (40 dní)

2) **Omyl konjunkcie [OK]:** "Predstavte si kocku so šiestimi stenami, ktorá má štyri steny zelené a dve červené. Kocku hodíme 20-krát po sebe a zapíšeme sekvenčiu zelených (Z) a červených (Č) stien, ktoré padnú. Predstavte si, že by ste mali vybrať spomedzi troch ponúkaných sekvenčí jednu, pričom by ste vyhrali 25 Eur, ak sa vybraná sekvencia objaví v rámci po sebe idúcich hodov kockou. Prosíme, zvoľte sekvenčiu zelených a červených, na ktorú by ste vsadili. a. ČZČČČ; b. ZČZČČČ; c. ZČČČČČ" (a)

3) **Finančná gramotnosť [FG]:** "Predstavte si, že vám banky ponúkajú uložiť si 3 000 eur na 10 rokov. Ktorú z týchto dvoch ponúk by ste si zvolili? a. vstupný poplatok 0,3%; priemerný výnos za 1 rok 3,0%; výstupný poplatok 0,5%; b. vstupný poplatok 1,0%; priemerný výnos za 1 rok 3,3%; výstupný poplatok 1,0%" (b)

4) **Numerická gramotnosť [NG]:** "Šanca vyhrat v lotérii je 1 z 10000. Aké percento ticketov, kúpených v rámci lotérie, by malo byť výherných? ____" (0,01)

5) **Kumulatívna pravdepodobnosť [KP]:** "Pre účel poistenia nehnuteľného majetku klienta proti následkom povodne je dôležité vyčísliť riziko výskytu povodní v danej lokalite. Pravdepodobnosť, že sa povodeň v určitej lokalite nevyskytne v priebehu jedného roka, je 90%. Aké je pravdepodobnosť, že sa povodeň v rovnakej lokalite nevyskytne v priebehu troch rokov? ____" ($72,9\% = 0,9 \cdot 0,9 \cdot 0,9 \cdot 100$)

6) **Problém Bertrandovej krabice [BK]:** "Predstavte si 3 urny. V jednej sú 2 zlaté mince. V druhej sú 2 strieborné mince. V tretej je 1 strieborná a 1 zlatá minca. Rozhodnete sa pre jednu z urien a zistíte, že ste si vytiahli zlatú mincu. Aká je pravdepodobnosť, že aj druhá minca v tejto urne je zlatá? ____" ($2/3 = 66,7\%$)

7) **Podmienená pravdepodobnosť** [PP]: "Reprezentatívny štatistický prieskum zistil, že 60% ľudí využíva bežný účet, ktorý nie je pre nich najvýhodnejší. Na trhu je aplikácia na vyhodnotenie, či je banka, ktorú človek využíva pre účel vedenia bežného účtu, preňho najvýhodnejšia. Aplikácia však nie je absolútne spoľahlivá. Zistilo sa, že ľudom, ktorí nevyužívajú najvýhodnejšiu možnosť, poradí zmenu banky v 90% prípadov. Ľudom, ktorí využívajú najvýhodnejšiu možnosť, v 5% prípadov však taktiež poradí, aby banku zmenili. Adam sa rozhodne vyskúšať túto aplikáciu a výsledkom je odporučenie, aby banku zmenil. Aká je pravdepodobnosť, že mu aplikácia poradila zle? ____" (2/54 = 3,7%)

Po vyriešení úloh participanti uviedli základné demografické údaje a zodpovedali otázku na apriórne iracionálne presvedčenia. Pýtali sme sa, či veria, že niektoré symboly a predmety prinášajú šťastie (1: vôbec neverím, 2: skôr verím, 3: neviem sa rozhodnúť, 4: skôr verím, 5: určite verím). Zameraním otázky na "symboly a predmety" sme sa pokúsili vyhnúť možným vplyvom inštrukcie o šťastnom čísle 7 v experimentálnej skupine na vyjadrenú mieru apriórneho iracionálneho presvedčenia.

3 Výsledky

Ako vidíme v Tab. 1, náročnosť úloh zodpovedala nášmu odhadu. Problém na omyl konjunkcie [OK] (pre ostatné skratky vid' Materiály a procedúra) však bol náročnejší než dve úlohy, ktoré po ňom nasledovali.

	KR	OK	FG	NG	KP	BK	PP
n	38	20	35	32	11	10	6
%	73,1	38,5	67,3	61,5	21,2	19,2	11,5

Tab. 1: Počet úspešných riešiteľov a percentuálna úspešnosť v jednotlivých úlohách

Celkové skóre sa pohybovalo od 0 do 7 a priemerná úspešnosť bola 2,9 ($SD = 2,0$). Odhady úspešnosti sa pohybovali medzi 0 a 7, no boli prevažne optimistické, keďže polovica ľudí uviedla predpokladané skóre 5 a viac ($Mdn = 5,5$; $IQR = 1,0$). Odhad a reálna úspešnosť spolu nesúviseli, $\rho = 0,03$; $p = 0,830$.

Čo sa týka predošej viery v šťastné symboly, odpovede sme kvôli nízkej početnosti v niektorých podskupinách rekódovali. Vznikli kategórie s nasledovným zastúpením: neverím ($n = 29$), neviem ($n = 6$), verím ($n = 17$).

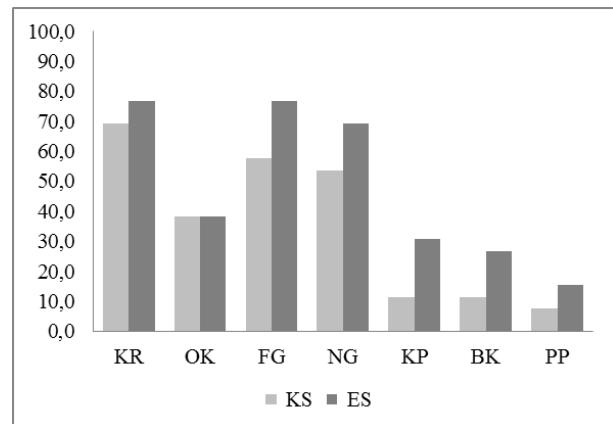
3.1 Rozdiely podľa rodu a veku

Medzi ženami a mužmi sme našli dva signifikantné rozdiely: ženy ($Mdn = 5,0$; $IQR = 2,0$) boli vo svojom odhade úspešnosti pesimistickejšie ako muži ($Mdn = 6,0$; $IQR = 2,0$), $M-W U = 168,5$; $p = 0,006$; $rm = 0,39$, a tiež sa od mužov líšili distribúciou apriórnych

presvedčení, $\chi^2(2) = 8,90$; $p = 0,012$; $V = 0,41$. Podiel mužov, neveriacich v šťastné symboly predstavoval totiž 70,4%, zatiaľ čo u žien to bolo iba 27,8%. Čo sa týka veku, participantov sme proporčne rozdelili na mladších (do 32 rokov) a starších (od 33 rokov), no žiadne významne rozdiely sme medzi nimi nenašli.

3.2 Vplyv navodeného iracionálneho presvedčenia

Porovnanie úspešnosti kontrolnej skupiny [KS] a experimentálnej skupiny [ES] v jednotlivých úlohách môžeme vidieť na Obr. 1 a v Tab.2. Hoci boli členovia ES o niečo úspešnejší v šiestich úlohách, žiadnen z rozdielov neboli štatisticky signifikantní.



Obr. 1: Úspešnosť v úlohách podľa skupiny

	KR	OK	FG	NG	KP	BK	PP
KS	69,2	38,5	57,7	53,8	11,5	11,5	7,7
ES	76,9	38,5	76,9	69,2	30,8	26,9	15,4
χ^2	0,39	0	2,19	1,30	2,88	1,98	0,75
p	,532	1	,139	,254	,090	,159	,385
ϕ	0,09	0	0,21	0,16	0,24	0,20	0,12

Tab. 2: Percentuálna úspešnosť v jednotlivých úlohách podľa skupiny a výsledky porovnania rozdielov

Skupiny sa významne nelíšili ani v celkovom skóre, kedy KS dosiahla priemer 2,5 ($SD = 1,9$) a priemer ES bol 3,4 ($SD = 2,1$), $t(50) = 1,53$; $p = 0,132$; $d = 0,42$. Členovia ES ($Mdn = 6,0$; $IQR = 2,0$) však odhadli svoju úspešnosť ako vyššiu než členovia KS ($Mdn = 5,0$; $IQR = 1,25$), $M-W U = 212,0$; $p = 0,016$; $rm = 0,33$. Zastúpenie kategórií apriórneho iracionálneho presvedčenia nebolo signifikantne odlišné, $\chi^2(2) = 3,59$; $p = 0,166$; $V = 0,26$, no podiel ľudí bez iracionálneho presvedčenia bol o niečo vyšší v ES (65,4%, $n = 17$) ako v KS (46,2%, $n = 12$).

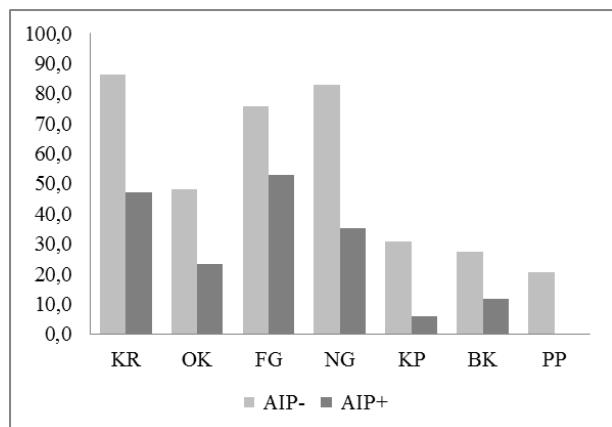
Vzhľadom na výsledky pilotnej štúdie sme sa osobitne pozreli na vplyv experimentálnej manipulácie u mužov a žien. V podskupine mužov sa kontrolná a experimentálna skupina zásadne nelíšili ani v odhade (KS: $Mdn = 5,0$; $IQR = 1,0$; ES: $Mdn = 6,0$; $IQR = 2,0$), $M-W U = 99,0$; $p = 0,137$; $rm = 0,28$, ani vo výkone

(KS: $M = 2,9$; $SD = 2,0$; ES: $M = 3,3$; $SD = 2,3$), $t(32) = -0,51$; $p = 0,614$; $d = 0,18$. U žien neboli rozdiel medzi kontrolnou a experimentálnou skupinou v odhadе úspešnosti štatisticky signifikantný, no jeho vecná významnosť bola vysoká (KS: $M = 4,6$; $SD = 1,2$; ES: $M = 5,4$; $SD = 0,8$), $t(16) = -1,70$; $p = 0,108$; $d = 0,82$. Podobne rozdiel v úspešnosti bol marginálne signifikantný, ale vecne veľký (KS: $M = 1,9$; $SD = 1,8$; ES: $M = 3,4$; $SD = 1,3$), $t(16) = -1,97$; $p = 0,066$; $d = 0,95$.

3.3 Rola apriórneho iracionálneho presvedčenia

V ďalšej časti analýzy sme sa zamerali na komparáciu skupín podľa pôvodného presvedčenia o šťastných symboloch. Pre tento účel sme vylúčili participantov s nerozhodnou odpoveďou. Výsledky porovnaní ľudí bez apriórneho iracionálneho presvedčenia [AIP-] a s apriórnym iracionálnym presvedčením [AIP+] prezentujeme na Obr. 2 a v Tab. 3.

Ludia, ktorí na šťastné symboly neveria boli významne úspešnejší v štyroch úlohách. Čo sa týka celkového výkonu, skupina AIP- ($M = 3,7$; $SD = 2,0$) skórovala výrazne lepšie než skupina AIP+ ($M = 1,8$; $SD = 1,6$), $t(44) = 3,42$; $p = 0,01$; $d = 1,04$, a tiež uviedla mierne vyšší odhad úspešnosti - rozdiel bol marginálne signifikantný (AIP-: $Mdn = 6,0$; $IQR = 1,5$; AIP+: $Mdn = 5,0$; $IQR = 1,5$), $M-W U = 169,5$; $p = 0,066$; $rm = 0,27$.



Obr. 2: Úspešnosť v úlohách podľa apriórneho presvedčenia

	KR	OK	FG	NG	KP	BK	PP
AIP-	86,2	48,3	75,9	82,8	31,0	27,6	20,7
AIP+	47,1	23,5	52,9	35,3	5,9	11,8	0,0
χ^2	8,10	2,76	2,56	10,6	3,99	1,58	4,05
P	,004	,097	,109	,001	,046	,209	,044
ϕ	0,42	0,25	0,24	0,48	0,29	0,19	0,30

Tab. 3: Percentuálna úspešnosť v jednotlivých úlohách podľa apriórneho presvedčenia a výsledky porovnania skupín

4 Diskusia

Kognitívna reflexia, numerická gramotnosť, porozumenie pravdepodobnostným konceptom a ďalšie kognitívne kompetencie, na ktoré sme sa zamerali v našom výskume, sú klúčovými prediktormi optimálnych úsudkov a rozhodnutí - či už v oblasti financií alebo i v iných doménach (e.g., Lipkus a spol., 2001; Navarrete a Mandel, 2015; Toplak a spol., 2011, 2014). Hoci sme v predošlých študiách zistili, že finanční profesionáli nepodliehajú niektorým kognitívnym odchýlkam, ako vplyv štatistického formátu (Dudeková a Kostovičová, 2016) alebo konverzačného kontextu (Kostovičová a Dudeková, 2016) na posudzovanie a rozhodovanie, úlohy v aktuálnom výskume patria medzi tie, kde je i v rámci vzdelanej a skúsenej populácie úspešnosť riešení nízka.

Polovici participantov sme sa snažili navodiť iracionálne presvedčenie, že by im číslo 7 mohlo priniesť šťastie pri riešení úloh. Nemôžeme si byť istí, či bola táto indukcia úspešná, keďže sme do dizajnu - podobne ako pôvodní autori - nezahrnuli kontrolu manipulácie. V porovnaní s aktiváciou presvedčení o šťastí typu priania "Držím vám palce" (Damisch a spol., 2010, Experiment 2) však nepredpokladáme, že by malo ísť o menej efektívnu manipuláciu.

Ak by navodenie, respektíve podpora, iracionálneho presvedčenia pomohla zlepšiť výkon v kognitívnych úlohách, bolo by to rozporuplné zistenie. Implikovalo by, že ide o vhodnú súčasť intervencií na reduciu kognitívnych odchýlok (tzv. "debiasing") a na podporu racionálnych rozhodnutí. Úlohy, ktoré zvolili originálni autori (Damisch a spol., 2010) boli viac zamenané na kognitívnu kapacitu (pamäť, generovanie slov), zatiaľ čo naše sa priamo týkali sklonov k chybám a (i)racionálneho prístupu k riešeniu problémov. Môže ísť o jeden z dôvodov, prečo sme efekt navodenia presvedčenia o šťastí, narození od pôvodnej štúdie, nenašli.

Ďalším vysvetlením je i samotná manipulácia. V rámci experimentov s kognitívnymi úlohami Damisch a spol. (2010) zvolili prítomnosť šťastného predmetu ako spôsob podpory iracionálneho presvedčenia a zlepšenia výkonu prostredníctvom zvýšenej seba-účinnosti. Ľudia si viac verili, stanovili si vyššie ciele v súvislosti s výkonom v úlohách a boli i výtrvalejší pri ich riešení, ak mali pri sebe šťastný predmet. Členovia našej experimentálnej skupiny boli sice tiež viac optimistickí, čo sa týka odhadu vlastnej úspešnosti, než participanti v kontrolnej skupine, tam však rozdiely medzi skupinami končia. Kým sme sa nepozreli na mužov a ženy osobitne. Podobne ako v pilotnej štúdii (Nedorostová, 2017) sme zistili, že aktivácia iracionálneho presvedčenia prospela len ženám. Zamýšľame sa, že rod ako moderátor efektu navodenia presvedčenia o šťastí môže súvisieť s výšou úzkosťou, nižšou sebadôverou a horším výkonom v logických a matematických úlohách u žien (Frederick, 2005; Meece a spol., 1990; Toplak a spol., 2014).

Okrem odlišných kultúr, v ktorých sa jednotlivé výskumy realizovali (Slovensko verzus Nemecko; Calin-Jageman a Caldwell, 2014; USA), sa naša vzorka líšila i profesionálnou skúsenosťou a najmä distribúciou apriórnych presvedčení. Experimentov, založených na prítomnosti šťastného predmetu sa, samozrejme, nezúčastnili ľudia, ktorí takýto predmet nevlastnia. Vzhľadom na menej početnú vzorku je zároveň možné, že boli skupiny viac heterogénne z hľadiska relevantných charakteristík typu štatistická gramotnosť či motivácia.

Naše hlavné zistenie sa týka rozdielov podľa pôvodného presvedčenia. Tí, čo v šťastné symboly a predmety verili, dopadli výrazne horšie vo väčšine kognitívnych úloh a aj v celkovom skóre. Zdá sa, že iracionálne presvedčenia, minimálne jedna ich kategória, neprospevia kognitívemu výkonu - či už numerickej gramotnosti, kognitívnej reflexii alebo posudzovaniu pravdepodobnosti. Presvedčenie, že niektoré symboly alebo predmety prinášajú šťastie, teda zodpovedá obsahu kontaminovaného mindware-u, ktorý je príčinou kognitívnych zlyhaní (Čavojová a spol., 2016; Stanovich, 2009). Príčinnosť vzťahov však v tomto prípade nie je možné určiť.

Suboptimálne rozhodnutia v oblasti financií môžu mať závažné individuálne i spoločenské dôsledky. Napríklad jednotlivci, ktorým sa nepodarí finančne sa pripraviť na dôchodok, budú predstavovať záťaž pre ekonomicky aktívnu populáciu. V prípade finančných profesionálov môžu ich rozhodnutiami byť ovplynení klienti spoliehajúci sa na ich odporúčania, či spoločnosti, ktorých prosperovanie závisí aj od dobrých finančných volieb. Práve preto je kapacita profesionálov robiť dobré finančné rozhodnutia ešte zásadnejšou než v bežnej populácii.

V súlade s predošlými výzvami (e.g., Čavojová a spol., 2016; Milkman a spol., 2009) by sme preto chceli zdôrazniť nevyhnutnosť systematickej kultivácie kritického myslenia a rationality, počnúc výučbou na školách, končiac rozvojovými programami pre zamestnancov a vedúcich pracovníkov. Intervencie na báze redukcie kognitívnych odchýlok a "opravy" kontaminovaného mindware-u by mohli ľuďom dopomôcť vyvarovať sa nežiaducích rozhodnutí v rámci profesionálneho a osobného života.

Poděkovanie

Príspevok vznikol za podpory Agentúry na podporu výskumu a vývoja na základe Zmluvy č. APVV-0361-12 - Rozhodovanie profesionálov: Procesuálne, osobnostné a sociálne aspekty

Literatúra

- Ariely, D. (2011). *Predictably irrational. The hidden forces that shape our decisions.* New York: HarperCollins Publishers.
- Aruguete, M. S., Goodboy, A. K., Jenkins, W. J., Mansson, D. H. a McCutcheon, L. E. (2012). Does religious faith improve test performance? *North American Journal of Psychology*, 14(1): 185–196.
- Bačová, V., Dudeková, K., Kostovičová, L. a Baláž, V. (2017). *Financial planning for retirement in young adults: Interaction of professional experience, knowledge, and beliefs.* Manuscript under review.
- Bar-Hillel, M. a Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, 11(2): 109–122.
- Čavojová, V., Ballová Mikušková, E., Majerník, M., Jurkovič, M., Juhásová, A. a Masaryk, R. (2016). *Rozum: návod na použitie. Psychológia racionálneho myšlenia.* Bratislava: IRIS.
- Calin-Jageman, R. J. a Caldwell, T. L. (2014). Replication of the Superstition and performance study by Damisch, Stoberock, and Mussweiler (2010). *Social Psychology*, 45(3): 239–245.
- Damisch, L. (2008). *Keep your fingers crossed! The influence of superstition on subsequent task performance and its mediating mechanism* (Doctoral dissertation). Cologne, Germany: University of Cologne.
- Damisch, L., Stoberock, B. a Mussweiler, T. (2010). Keep your fingers crossed! How superstition improves performance. *Psychological Science*, 21(7): 1014–1020.
- Dudeková, K. a Kostovičová, L. (2016). Čo z čoho a kto z koho? Efekt číselného vyjadrenia a odbornosť pri posudzovaní rizika v oblastiach zdravia a financií. V zborníku *Kognice a umělý život XVI*, str. 23–28.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4): 25–42.
- Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.
- Kostovičová, L. a Dudeková, K. (2016). Finančníci – pragmatici? Konverzačný kontext a doménová špecifickosť v posudzovaní rizika a rozhodovaní. V zborníku *Kognice a umělý život XVI*, str. 85–89.
- Lee, C., Linkenauger, S. A., Bakdash, J. Z., Joy-Gaba, J. A. a Profitt, D. R. (2011). Putting like a pro: The

role of positive contagion in golf performance and perception. *PloS One*, 6(10): e26016.

Lipkus, I. M., Samsa, G. a Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1): 37–44.

McClay, R. a Beaman, C.P. (2005). Problem-structure and format in training conditional and cumulative risk judgements. V zborníku *27th Annual Conference of the Cognitive Science Society*, str. 1449–1454.

Meece, J. L., Wigfield, A. a Eccles, J. S. (1990). Predictors of math anxiety and its consequences for young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, 82(1): 60–70.

Milkman, K. L., Chugh, D. a Bazerman, M. H. (2009). How can decision making be improved? *Perspectives on Psychological Science*, 4(4): 379–383.

Navarrete, G. a Mandel, D. R. (Eds.) (2016). *Improving Bayesian reasoning: What works and why?* Frontiers Media SA.

Nedorostová, R. (2017). *Efekt šťastných ponožiek: Zvýšenie úspešnosti riešení úloh prostredníctvom podpory iracionálnych presvedčení?* Manuscript in preparation.

Stanovich, K. E. (2009). *What intelligence tests miss. The psychology of rational thought.* New Haven, CT: Yale University Press.

Stanovich, K. E. (2011). *Rationality & reflective mind.* New York: Oxford University Press.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. V knihe *The Oxford handbook of thinking and reasoning*, str. 343–365.

Toplak, M. E., West, R. F. a Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7): 1275–1289.

Toplak, M. E., West, R. F. a Stanovich, K. E. (2014). Assessing miserly processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2): 147–168.

Tversky, A. a Kahneman, D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4): 293–315.

Graphical Granger Causality for Detection Temporal Anomalies in EEG Data

Anton Kováč¹, Katerína Hlaváčková-Schindler²

(¹) Faculty of Mathematics, Physics, and Informatics, Comenius University in Bratislava,

Mlynská dolina, 84248 Bratislava, Email: tondo.kovac@gmail.com

(²) University of Vienna, Währinger Straße 29, 1090 Vienna

Abstract

The goal of our work was to apply the algorithm of (Qiu et al., 2012), using Graphical Granger causality for time series anomaly detection so that it can be used to the EEG time series of a human brain. The authors presented the algorithm in a brief form and its parametrization is not sufficiently discussed. We elaborated its detailed parametrization and focused on its practical utilization for detection of anomalies in EEG data, particularly for sleep EEG data analysis. For such purpose we created the program in Matlab software.

Our work was intended mainly for introduction of our program to interested users. It may offer rigorous testing of EEG data in terms of temporal anomaly discovery in the brain connectivity. Our program provides a relatively detailed view at the temporal and spatial domain of brain activity. Nevertheless, the program should be considered as the beta version and we are still working on the details.

1 Introduction

The Graphical Granger models include family of graphical models that exploit temporal dependencies between a larger number of variables by applying Granger causality paradigm. By definition, Granger causality is a statistical concept of prediction according which the signal X “Granger causes” signal Y if the past values of X can help to build a better model for prediction of Y with respect to the model with only the past values of Y (Granger, 1969; Cohen, 2014; Seth et al., 2015). While the concept of Granger causality is bivariate (i.e. it describes the relationship between two time series), the later introduced Graphical Granger approach is designed for multivariable time series data (Arnold et al., 2007). Arnold et al. (2007) applied L_1 penalized regression to multivariate causality testing which enabled them to avoid a multiple statistical hypothesis testing, connected with the bivariate Granger causality.

In our work we selected EEG data recorded on a stroke patient during sleep. The settings in sleep can relatively reliably control the contamination of basal brain activity or random thought processes which may occur during the specific testing (e.g. focus on ERP, or task related brain activity).

The research of the stroke patients during sleep has shown that the stroke patients can suffer from sleep disorders (Vock et al., 2002; Hermann et al., 2008). Further, in Müller et al. (2002) the authors found a lower duration of the REM and the slow-wave sleep phase; In Baumann et al. (2006) an increased duration of NREM sleep phase in stroke patients was found compared to the control group. However, persisting sleep abnormalities are rather unusual (Vock et al., 2002). Nevertheless, such conclusion may emerge from rather poor methods for detecting specific abnormalities. To our best knowledge, the research of either the time duration of these anomalies or their concrete localization has not been done yet.

In our work we focused on the detection of temporal anomalies. The idea of Qiu et al. (2012) provides a promising tool to study time dependent anomalies in EEG among various brain areas.

We created the program in Matlab software which applies the basic algorithm designed in Qiu et al. (2012). The program takes the EEG data files (the raw signal and information about the possible artifacts identified preliminarily) and tests the model for different parameters. The output of the program is the graph where possible anomalies are highlighted. However, the program should be considered as the beta version and we are still working on the details.

2 Methods and data

The basic algorithm provided by Qiu et al. (2012) for temporal anomaly detection consists of three steps:

1. Learning temporal causal graph of test set by regularization
2. Computing the anomaly scores of test set using Kullback-Leibler divergence
3. Determining anomaly by threshold cutoff and identifying potential causes.

In the following we explain the whole algorithm in detail.

(Step 1) We selected a reference set and a specific test set where the anomalies will be examined. We use

the same notation as in Qiu et al. (2012):

$$\begin{aligned} D^{(a)} &= \{x_i^{(a),t} | i = 1, \dots, p; t = 1, \dots, T^{(a)}\} \\ D^{(b)} &= \{x_i^{(b),t} | i = 1, \dots, p; t = 1, \dots, T^{(b)}\} \end{aligned} \quad (1)$$

where T is the length of the time series, p is the dimension (the number of features respectively), $D^{(a)}$ is the notation for reference set, and $D^{(b)}$ is the notation for test set.

(Step 2) We compute the regression coefficients of reference set using regularized L_1 regression based on Graphical Granger approach (Qiu et al., 2012). This step is described by the following equation

$$\hat{\beta}_i(\lambda) = \arg \min_{\beta_i} \left(\sum_{t=1}^T \|x_i^t - X_{t,L}^{\text{Lagged}} \beta_i\|^2 + \lambda \|\beta_i\|_1 \right) \quad (2)$$

where $X_{t,L}^{\text{Lagged}}$ represents the concatenated vector of all the lagged variables (with maximal lag of L) up to time t , i.e. $x_i^{t-l} : j = 1, \dots, p; l = 1, \dots, L$, $L > 0$, and λ is the shrinkage parameter in the LASSO algorithm, i.e. $\lambda > 0$. We mention that the multivariable time series X_i^t is given in the form $X_i^t : t = 1, \dots, T; i = 1, \dots, p$.

Next, we are interested in computing the anomaly score for $D^{(b)}$ with respect to $D^{(a)}$, whether and how much each variable contributes to the difference between these two datasets.

(Step 3) We compute regression coefficients for $D^{(b)}$ using learning temporal causal graph by regularization. The authors in Qiu et al. (2012) consider the null hypothesis of no temporal anomaly which is that the temporal causal graphs for reference and test set are the same. Thus, we can use the null hypothesis as an additional constraint in the optimization algorithm outlined in Granger LASSO algorithm. We choose the coefficient similarity regularization term, which is the form

$$\hat{\beta}_i^{(b)}(\lambda) = \arg \min_{\beta_i} \left(\sum_{t=1}^T \|x_i^{(b),t} - X_{t,L}^{(b),\text{Lagged}} \beta_i^{(b)}\|^2 \right) \quad (3)$$

subject to

$$\sum_j |\beta_{i,j}^{(b)} - \beta_{i,j}^{(a)}| \leq \epsilon$$

The problem in equation (3) corresponds to the optimization problem (4) in Qiu et al. (2012).

(Step 4) We compute the anomaly score using Kullback-Leibler divergence. We assume that the underlying distributions are normal (thanks to the central limit theorem for large data sets) and this allows us to use the equation (6) from Qiu et al. (2012):

$$d_i^{(ab)} = -(ln|\sigma^{(a)}| - ln|\sigma^{(b)}|) - \frac{1}{2} \left(1 - \frac{1}{(\sigma^{(b)})^2} ((\sigma^{(a)})^2 + ((\beta_i^{(a)} - \beta_i^{(b)})\Delta)^2) \right) \quad (4)$$

where $(\beta_i^{(a)}, (\sigma^{(a)})^2$ and $(\beta_i^{(b)}, (\sigma^{(b)})^2$ are coefficients and variance for $D^{(a)}$ and $D^{(b)}$ respectively, and Δ is the mean of the concatenated lagged variables.

These divergence score is then computed in the reverse direction (d^{ab} and d^{ba}). Qiu et al. (2012) define the anomaly score of i -th variable as:

$$a_i \equiv \max \{d_i^{ab}, d_i^{ba}\} \quad (5)$$

This approach allows us to slide a window of a specific length to test and compute the anomaly score within this window and thus identify a potential time anomaly. Then we compute the threshold cutoff for determine what the anomaly is and what is not. We slide the time windows of the reference set and compute the anomaly score for each window. We use this process to approximate the distribution of the anomaly scores, which normal time series should have. Given a significance level of α , we use α -quantile of this distribution as the threshold cutoff. Therefore, this approach allows not only to determine, if the whole time series is anomalous, but also in which particular time windows the anomaly occurs.

In our work we focused to investigate the algorithm with respect to the parameters:

- λ – penalty constant of LASSO Granger
- L – model order or maximum time lag
- ϵ - constraint parameter.

3 Results and outlines

In the experiments we tested the performance of our program with different samples of reference and test set and parameter settings. We tested the program on the sleep EEG data from a male subject after a stroke. The signal was recorded on 6 electrodes ('Fp1M2' - 'Fp2M1', 'C3M2' - 'C4M1' and 'O1M2' - 'O2M1' pairs) on 200Hz resolution. The data was recorded during the whole two nights. The raw signal was cleaned from any artifacts (we considered only the clear signal without any artifacts). The sleep structure was analyzed in 30 s epochs according to the standards given from Rechtschaffen and Kales, the scoring rules for the sleep (Rechtschaffen a Kales, 1968). Thus, we had also the information about the sleep stage in which the subject was occurred. Finally, we used 95 quantile of the anomaly distribution in reference data to determine the threshold cutoff. We did not perform any other transformation of the data. We followed the basic instructions to analyze the neural time series data (Cohen, 2014). These are:

- We interpret an anomaly detected by the program as an abnormality in the functional brain connectivity.

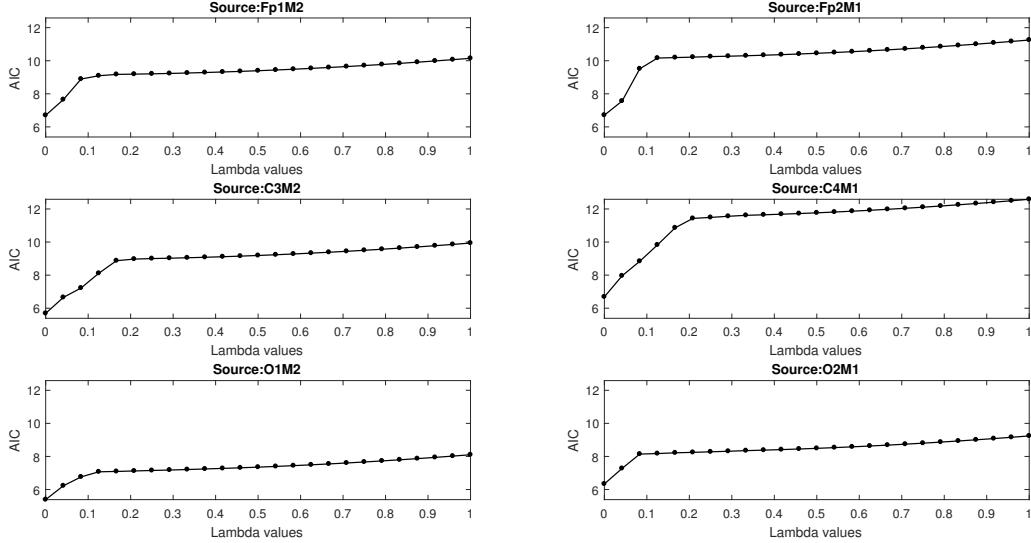


Fig. 1: Comparison of the AIC for different values of lambda parameter.

- We performed our analysis on the raw EEG signal.
- We focused on shorter time windows (100s for reference set, 80s for test set) to avoid non-stationarity occurrences in the data.

In the following section our program and its achieved outputs with respect to the mentioned parameters are presented. We picked the outputs from the testing anomaly in S4 stage according to Rechtschaffen a Kales (1968) scoring. We set for the reference data second S4 stage recorded during first sleep record and for the test set third S4 stage from second sleep record. Since there is an evidence of prolonged NREM phase (Baumann et al., 2006), we focused on the anomaly close to the end of the phase. Such settings can provide an information about the location of some unspecific processes represented in detected anomaly.

3.1 Testing the parameter λ

Since the adopted algorithm from Qiu et al. (2012) requires several user defined parameters, we tried to find their optimal values. We fit 25 models on the reference data with the λ parameter uniformly distributed in interval $(0, 1)$. The Fig. 1 illustrates the evaluation of fitted models based on Akaike Information Criteria (AIC).

In the Fig. 1 we can see the AIC for all nodes in a graph (electrodes) with different settings of parameter λ . In general, the major increase of the AIC is seen up to value of λ set to 0.2 (in electrode “C4-M1”). The models after that value show slightly increasing of AIC. According to AIC of all the models we set the lambda parameter to value 0.15. Note that the lowest AIC was on the initial lowest value of lambda. However, this parameter refers to the value of the penalty coefficient in

LASSO Granger model. Therefore it is reasonable to set it to the slightly higher value, to increase the importance of the penalty function for the whole AIC.

3.2 Testing the model order L

As in the previous part we tested the possible model order (different maximum lag L) in terms of LASSO Granger on the reference set. We set the maximum lag L from 1 up to 10. The Fig. 2 illustrates the output of this testing.

The models with different time lag L parameter show the biggest AIC improvement with time lag L from 4 to 7. Nevertheless, adding the regressors may lead to explain more variability in the targeting feature, however it may also lead to a higher risk of multicollinearity. Therefore, according to AIC of all the models we performed our model experiment with time lag $L = 4$.

3.3 Testing the parameter ϵ

In our experiments we tested the constant in the constraint from equation (3) with discrete values ($\epsilon = \{0.01, 0.1, 1, 10, 100, 500\}$). In most cases the @fmincon solver from MATLAB optimization toolbox found optimal solutions for equation (3) with constant $\epsilon = 10$. However, this constant needs more attention in the future research. The original article (Qiu et al., 2012) did not give any hint how to select this parameter.

3.4 Interpretation of the detected anomalies

The final output of our program shows the signal from particular electrodes in the user defined time window. Thus, the user can select a specific sleep phase and ex-

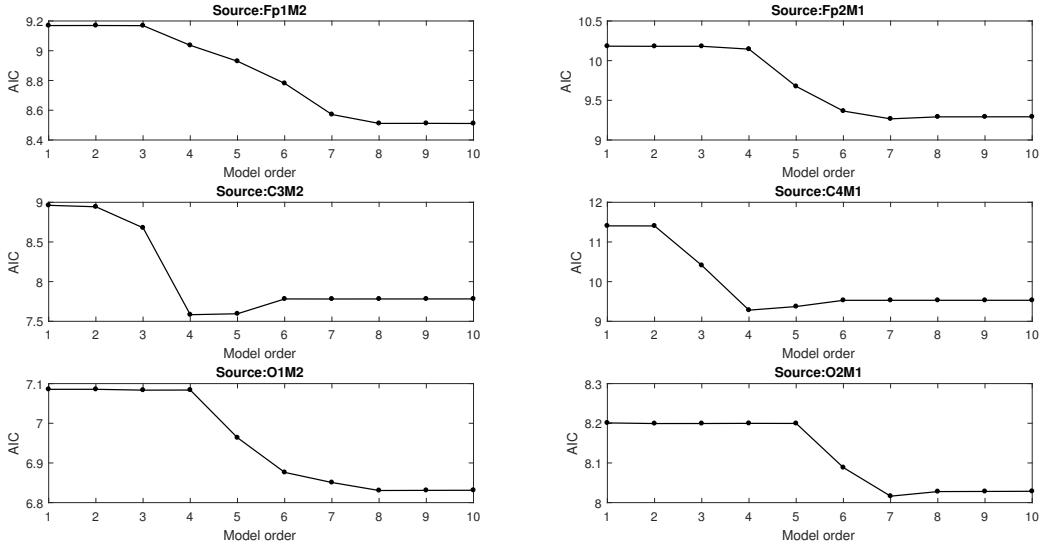


Fig. 2: Comparison of the AIC for different maximum lag.

amine the anomaly in the test set defined by user.

In the Fig. 3 we can see the signal from all electrodes. The highlighted areas represent identified anomalies on the one second time window scale. The advantage of our approach is that we can specify various time segments in the test data and examine them for the anomaly.

4 Conclusion and outlook

In our data we did not find any suspicious anomalies in specific sleep stages. However, we can interpret the anomalies which we detected as the abnormalities in the brain areas under the particular electrodes. In our presented example we can see that the anomalies occur rather in central and dorsal area of the brain. As it was mentioned earlier, the previous research of stroke patients during the sleep has not shown any persisting abnormalities (Vock et al., 2002). Such claim is consistent with our findings of relatively small number of anomalies. Furthermore, our method provides a detailed view on the EEG data and thus it can lead to create new hypotheses in the neuroscience research, specifically, in our case, in the research of the sleep of the patients after a stroke. For instance, regarding to our presented example we can assume that our tested patient may suffer to a stroke related symptoms that lead to abnormal functioning in the central and dorsal area of the brain. However, the more experiments and validation is required in this area. E.g. we suggest the comparison of the healthy control subjects with the patients after a stroke. We can also test other sleep phases. Furthermore, to standardize and parametrize our program we can perform more experiments on the healthy population to see optimal

parameter settings for EEG sleep analysis.

This outline was intended mainly for introduction of our program to interested users. We suggest that our program may offer rigorous testing of EEG data in terms of finding the anomalies in the brain connectivity. The advantage of our program is in the relatively detailed view at the temporal and spatial domain of the brain activity. The user can specify time scale to be explored, even to short time windows. However this may lead to some drawbacks, as it increases the possibility of non-stationarity of the signal.

Our future goal is to further parametrize and validate approach that we proposed in this work. We need to test the algorithm on the more data and to standardize parameter settings for specific purposes. Further, it seems necessary to discuss the various problems with experts in particular fields to compose exact definition of what a temporal anomaly really is and what interpretations seem reasonable. Furthermore, our implementation of algorithm provided in Qiu et al. (2012) is (at least to our knowledge) the first attempt to detect temporal anomalies in this data and settings (sleep EEG data, patient after stroke). Therefore, this work may be considered as the first introduction of approach to examine sleep EEG time series data.

Finally, we are aware that this work still does not provide optimal parametrization but it should be considered as the beginning of the research work.

Acknowledgment

Our special thanks go to Dr. Roman Rosipal from Slovak Academy of Science for providing the data for testing our implementation of the algorithm. Further, our

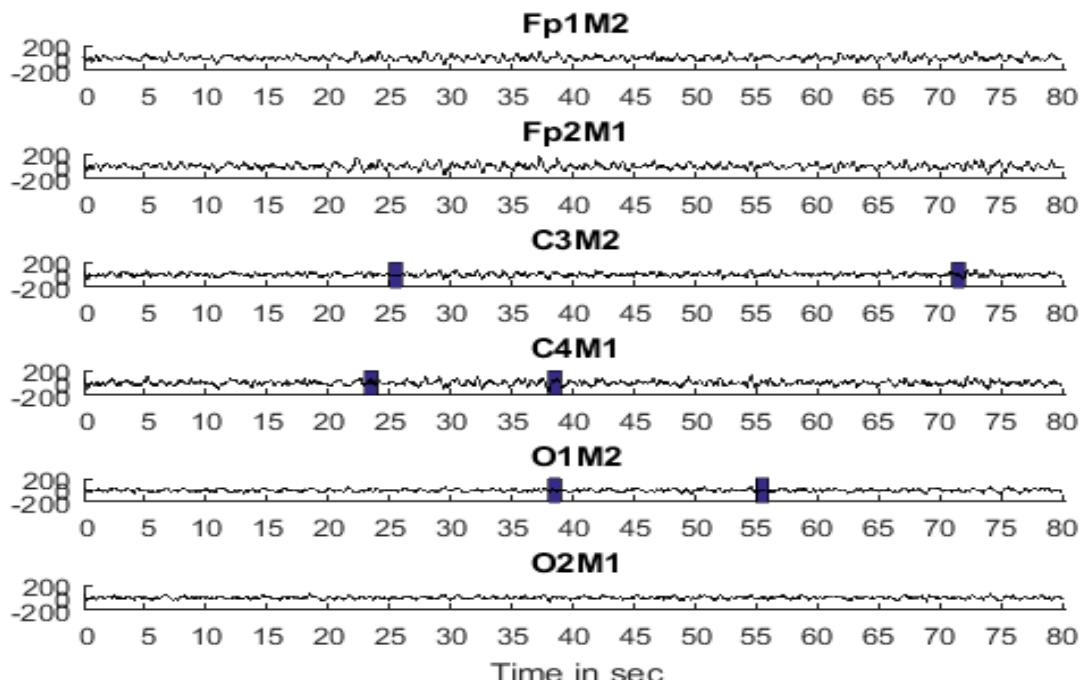


Fig. 3: EEG signal and highlighted anomaly.

thank goes to researchers in Institute of Measurement in Slovak Academy of Sciences, namely Mgr. Zuzana Rostakova, Dr. Anna Kravcovska and Dr. Martina Chvostekova for valuable discussion and recommendation in analysis.

References

- Arnold, A., Liu, Y. a Abe, N. (2007). Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 66–75.
- Baumann, C. R., Kilic, E., Petit, B., Werth, E., Hermann, D. M., Tafti, M. a Bassetti, C. L. (2006). Sleep eeg changes after middle cerebral artery infarcts in mice: different effects of striatal and cortical lesions. *Sleep-New York Then Westchester-*, 29(10):1339.
- Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pp. 424–438.
- Hermann, D. M., Siccoli, M., Brugger, P., Wachter, K., Mathis, J., Achermann, P. a Bassetti, C. L. (2008). Evolution of neurological, neuropsychological and sleep-wake disturbances after paramedian thalamic stroke. *Stroke*, 39(1):62–68.
- Müller, C., Achermann, P., Bischof, M., Nirko, A. C., Roth, C. a Bassetti, C. L. (2002). Visual and spectral analysis of sleep EEG in acute hemispheric stroke. *European neurology*, 48(3):164–171.
- Qiu, H., Liu, Y., Subrahmanyam, N. A. a Li, W. (2012). Granger causality for time-series anomaly detection. In *ICDM, 2012 IEEE 12th International Conference on Data Mining*, pp. 1074–1079.
- Rechtschaffen, A. a Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. US Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, National Institute of Neurological Diseases and Blindness, Neurological Information Network.
- Schmidt, M. (2005). Least squares optimization with 11-norm regularization. *CS542B Project Report*, pp. 14–18.
- Seth, A. K., Barrett, A. B. a Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297.
- Vock, J., Achermann, P., Bischof, M., Milanova, M., Müller, C., Nirko, A., Roth, C. a Bassetti, C. L. (2002). Evolution of sleep and sleep EEG after hemispheric stroke. *Journal of Sleep Research*, 11(4):331–338.

Robotic reaching and grasping using a TRPO-based reinforcement learning

Peter Kovács, Igor Farkaš

Faculty of Mathematics, Physics and Informatics

Comenius University in Bratislava

Email: kovacs.pjotr@gmail.com, farkas@fmph.uniba.sk

Abstract

We have designed a simulation of a robotic arm for object reaching and grasping using reinforcement learning approach and trained it with recently introduced TRPO - Trust Region Policy Optimisation algorithm (Schulman, 2015) by controlling forces applied to each degree of freedom. We also implemented a parallel TRPO to speed up the training procedure and introduced improvements regarding the sample efficiency. Finally we evaluated TRPO algorithm in different extended scenarios such as grasping a randomly placed target or reaching for the target behind the wall.

1 Introduction

Finding a suitable grasp among infinite set of candidates is a challenging task that has been addressed frequently in the robotics community resulting in different approaches such as in Kraft et. al. (2010), Castellini et. al. (2007), Kalakrishnan et. al. (2011) and Zdechovan (2012). The methodologies can be divided into two categories analytic solutions using methods of inverse kinematics/dynamics and data driven approaches (Bohg et. al., 2014). Until recently the field of robotic grasping was clearly dominated by analytic approaches. However, with development of fast physics simulators and new advances in the field of machine learning and neural networks research lead to new data driven approaches.

Data driven approaches can be divided into three categories. The first category covers methods that assume knowledge of 3D mesh and the challenge is then to sample a set of good grasp hypotheses and rank them according to some quality measure. The second approach would be learning from humans when robot can learn successful grasps by observing them. This method is often referred to as demonstration or imitation learning. The last category is learning through trial and error. Instead of computing possible grasps we just try multiple of them and learn some generalized knowledge from such experience. Reinforcement learning is a good representative of this category.

Kalakrishnan et. al.. (2011) showed successful application of RL to door opening and pen grasping task. Zdechovan (2012) showed how to use CACLA (van Hasselt Wiering, 2009) to learn successfully grasp

object of random shapes with an actor–critic algorithm. Lillicrap et. al. (2015) showed that its also possible to employ a deep learning with actor–critic methods to learn good models for various robotic tasks. Recently Schulman (2015) proposed TRPO algorithm which showed promising results on robotic locomotion and motivated us to test this algorithm on robotic reaching and grasping task.

Formal definition of reinforcement learning comes from Markov decision process (MDP) theory defined by the tuple $(S, A, P, r, \rho_0, \gamma)$. where S is a set of states. (e.g., in robotic arm grasping, S might be positions and angles of joints and joint torques), A is a set of actions. (e.g. all possible changes in torque joints), $P : S \times A \times S \rightarrow R$ is the transition probability distribution (model), $r : S \times A \times S \rightarrow R$ is the reward function, $\rho_0 : S \rightarrow R$ is the distribution of the initial state s_0 and $\gamma \in (0, 1)$ is the discount factor. Let π denote a stochastic policy $\pi : S \times A \rightarrow [0, 1]$ and $\eta(\pi)$ denote its expected discounted reward: $\eta(\pi) = \mathbb{E}_{s_0, a_0, \dots} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t) \right]$ where $s_0 \sim \rho_0(s_0)$, $a_t \sim \pi(a_t | s_t)$, $s_{t+1} \sim P(s_{t+1} | s_t, a_t)$.

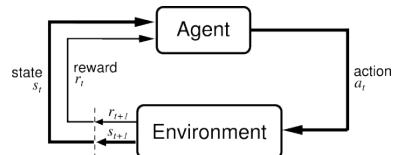


Figure 1: RL dynamics

The goal of RL is to find a policy π that it maximizes $\eta(\pi)$. MDP dynamics is shown in Figure 1. The agent observes current world state s_t , and performs action a_t , while receiving reward r_t and transforming the world into a new state s_{t+1} .

2 Trust region policy optimization

The problem can be formally formulated as optimisation problem in the following way:

$$\begin{aligned} \max_{\theta} \quad & \mathbb{E}_{s \sim \rho_{\theta_{old}}, a \sim q} \left[\frac{\pi_{\theta}(a|s)}{q(a|s)} Q_{\theta_{old}}(s, a) \right] \\ \text{s.t.} \quad & \mathbb{E}_{s \sim \rho_{\theta_{old}}} [D(\pi_{\theta_{old}}(\cdot|s) \| \pi_{\theta}(\cdot|s))] \leq \delta \end{aligned} \quad (1)$$

where $\frac{\pi_\theta(a|s)}{q(a|s)} Q_{\theta_{old}}(s, a)$ is unbiased cumulative reward of new policy $\pi_\theta(a|s)$, $q(a|s)$ is our sampling scheme, $\frac{\pi_\theta(a|s)}{q(a|s)}$ is importance sampling ratio and $Q_{\theta_{old}}(s, a)$ is an action-state value function approximator. D is an KL-divergence of two probability distributions, in this case our stochastic policies and δ is hyper-parameter of our model which describes how much two policies can differ between each update.

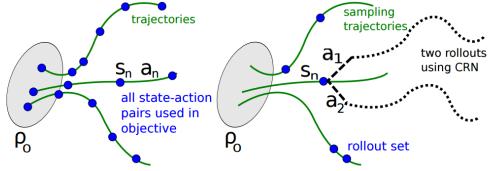


Figure 2: Demonstration of single path sampling scheme (left). Vine sampling procedure (right). (taken from Schulman (2015))

In TRPO Schulman (2015) describes two different sampling schemes single path and vine. In single path we sample multiple states and then follow policy $\pi_{\theta_{old}}$ for some number of time-steps to generate trajectories. In vine method we would additionally select multiple actions in multiple states to better approximate cumulative reward. However this would require simulator to be restored to particular states. In our implementation we decided to focus on simple path sampling scheme due to its better computational efficiency and easier implementation.

Algorithm: TRPO

- 1: Initialize policy parameter θ
 - 2: **while** TRUE **do**
 - 3: Use sampling scheme to collect set of state-action pairs with cumulative rewards.
 - 4: By averaging over samples, construct the estimated objective and constraint.
 - 5: Compute search direction using linear approximation to objective and quadratic to constraint.
 - 6: Perform line search in that direction to improve $\mathbb{E}_{s \sim \rho_{\theta_{old}}, a \sim q} \left[\frac{\pi_\theta(a|s)}{q(a|s)} Q_{\theta_{old}}(s, a) \right]$ and satisfy the constraint on δ .
 - 7: **end while**
-

3 Model specifications

3.1 State representation

In our simulated world the state is represented by vector of positions for each degree of freedom (DoF) followed by position of the target and actual forces/torques on each DoF. After observing current state and reward for

last step agent provides 1 action for each degree of freedom. This action is represented as force/torque applied to each degree of freedom respectively.

3.2 Policy

As our stochastic policy we used fully connected artificial neural network, which means that each input neuron is connected with each output neuron in the following layer. In order to scale on each environment we have taken relatively big network consisting of two hidden layers with 100 neurons in each. As a nonlinear function we have used hyperbolic tangent. And as initialization method we used Xavier initialization Glorot Bengio (2010) which helps training when we have more layers.

We have experimented with activation functions such as ReLU variations and sigmoidal nonlinearities. However we found that TRPO is quite invariant to these changes and converged almost in all settings with minor improvements in some combinations, therefore we decided to stick with standard tanh activation function. Our experiments with number of layers and neurons in each layer showed that using smaller networks improved convergence in easier tasks such as 2D reaching but was for example not sufficient in 3D grasping. Therefore we decided to stick with already mentioned architecture which provides more than enough parameters for learning grasping policy.

3.3 Value function

As Q function approximator we have stucked with simple linear regression. For training we have used closed form for least squares. However when trying to learn policy we have discovered that we are unable to learn due any meaningful value function approximation due to non-linearity of underlying value function. We partially solved this by nonlinearizing input to regression by giving it also squares of each input dimensions.

3.4 TRPO parameters

In TRPO the most sensitive hyperparameters are maximum KL divergence and the number of rollouts per update. This is probably intuitive since the whole algorithm convergence properties depend on these two. In all of our experiments we use maximum KL divergence $\delta = 0.01$. Regarding number of sampled trajectories we increase it with difficulty of the environment.

4 Experiments

4.1 3D Reaching

In 3D reaching environment we have placed rigid ceiling on which our arm will hang. The whole arm have four DoF which enables it to reach almost all positions in reaching area bounded by red walls (Figure 3). The goal of this environment is to reach for the red(target)

with the green dot(palm). At the beginning of each reaching the target is moved to random location and random small force is applied to each of the arms degree of freedom.

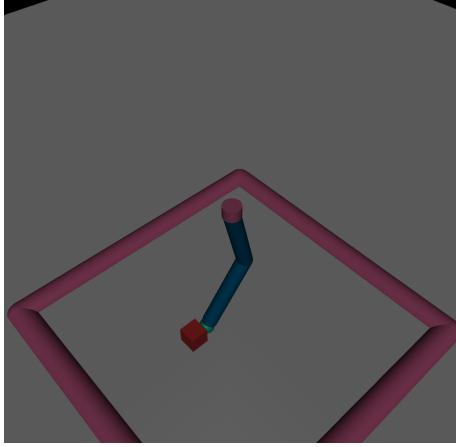


Figure 3: Rendered Reacher3D environment

After each step environment provides agent with two different information - state and reward. Based on this information agent select actions. As state agent is given angles for each DoF, position of target in Cartesian coordinate system and current velocity of each DoF. As reward agent is given

$$r = -d(\text{palm}, \text{target}) - f(\text{actions}) \quad (2)$$

1. Distance $d(\text{palm}, \text{target})$ of palm from target - The higher the distance the lower the reward is. This forces algorithm to find policy which tries to have green and red dot as near as possible.
2. Moving penalty $f(\text{actions})$ - We penalize robot for applying force on joints. It ensures that robot try to find optimal way to reach his target and do minimal work.

As we can see (Figure 4 such problem formulation allows our algorithm to successfully converge to good policy for reaching on randomly placed target. Policy controlling the robotic arm can also be viewed on video¹.

4.2 3D Grasping

For the extension of 3D reaching environment to 3D grasping we added 4 fingers where each side is coupled together and is controlled by one degree of freedom. To generate valid grasps we also added rotational degree of freedom on forearm of our robotic arm. Together this means we are now controlling 7 degrees of freedom. After multiple environments we noticed that having 2 degrees of freedom for each part of claw is unreasonable and we decided to constraint them together to have same inverse positions to each other. This means if left

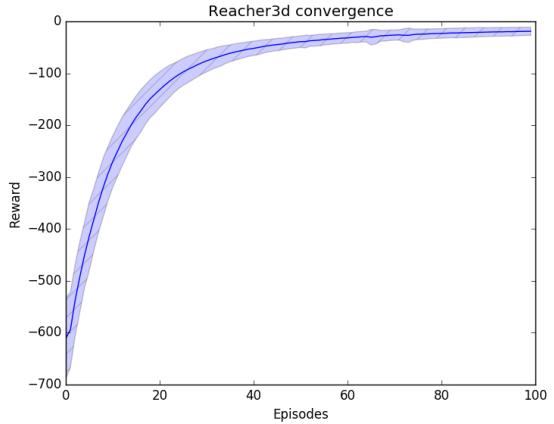


Figure 4: A cumulative reward during training with its uncertainty over 10 runs.

part of claw is at 50 degrees than right part must be at -50.

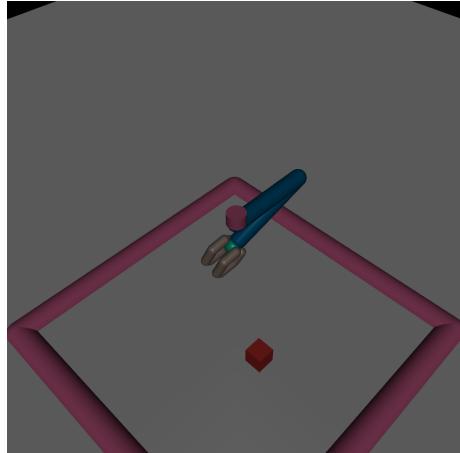


Figure 5: Rendered Reacher3D environment

Reward function in this environment is a bit more complicated than in the reaching part. It still have same two parts as in 2 but with some additional parts which force arm to grasp target. The first such part is positive reward for touching claws with object. Second part is penalty for contact between ground and claw. Reasoning behind this is that if we are penalized for ground contact we will try to lift the object from ground. The last part of reward is for the height of the target.

The second part forces the arm not to touch ground when we have successfully grasped the target and therefore arm will lift the object leading to big positive reward for target height. Results of our training are visible in either graph 6 or in video² from training.

¹https://youtu.be/_8uSyV750tE

²<https://youtu.be/EGvkJBXAOi4>

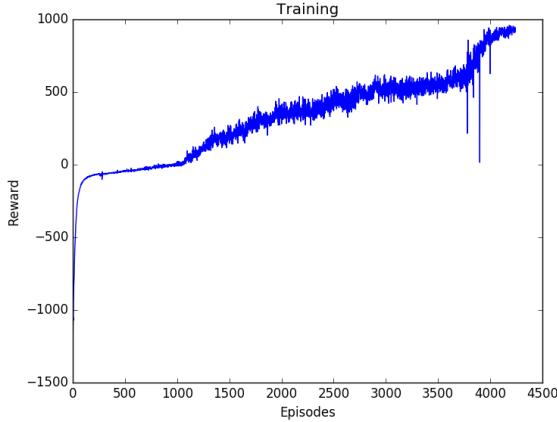


Figure 6: A cumulative reward during training.

4.3 Reaching behind the wall

Simple grasping as shown in previous section is really useful for various robotic tasks, but for robots operating in real world there are often different obstacles on the way to target which needs to be avoided so robot or obstacle is not destroyed. Hence we have designed another simple environment (Figure 7) consisting of reaching area, 6 DoF arm and wall in the middle of reaching area.

At the start of the episode target is spawned on random position in reaching area but not on the wall. Then robotic arm is moved to the opposite side of the wall as its target and its goal is to reach for that target without touching the wall. Policies using former reward function (2) were unable to find policy which avoided the wall, thus we have again modified it. The modification penalized each contact with wall and also penalized arm for being closer to wall than to target. This lead to desired behaviour of arm going over the wall and then reach for target. Video³ from training is available on youtube.

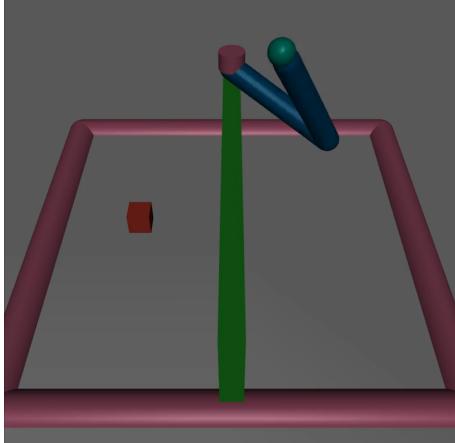


Figure 7: Rendered Grasper3D_wall environment

³<https://youtu.be/dXAFzf1knDI>

5 TRPO improvements

We propose a few TRPO improvements we have tried in order to speed up computation time or convergence rate. Firstly we will introduce experience replay (Lin, 1992) and experiment on improving sample efficiency by using idea based on experience replay. Secondly we will look into reusing previous gradient direction information by using it as initialization in conjugate gradient algorithm. In the end we will describe our parallel version of TRPO algorithm.

5.1 Replay memory

After performing multiple experiments we have noticed that training TRPO is really time consuming. We were wondering if we could somehow speedup network convergence properties. One of the first ideas was to use collected trajectories not only in current iteration but also in the future. Specifically we have decided to keep trajectories from past k steps.

The way to reuse them was to use past samples with correctly-set importance weight. However this meant additional computation of $Q_{\theta_{old}}(s, a)$ also for trajectories from past iterations. This idea led to minor improvement on each tested environment as can be observed on Figure 8 .

Experiment was conducted on 10 runs with different random seed for each alternative. The dark blue/green line indicates the average of average sum of rewards per episode while a transparently colored surrounding indicates the uncertainty. In each episode there were sampled 100 trajectories and replay memory hyperparameter was set to $k = 3$. As we can see from graphs this technique reduces variance of rewards and also slightly improves performance of original algorithm consistently across multiple environments. It seems that the more complex the environment is the more this method helps. Such statement would require further investigation which we could not afford due to time/computational power constraints.

5.2 Using previous direction in CG

Another possibility of reusing previous information is based on reusing previous gradient instead of reusing experience explicitly as was replay memory case. As was stated in Martens (2010) a simple enhancement to HF algorithm which they found improves performance by an order of magnitude was to reuse previous search direction found by conjugate gradient (CG) algorithm and use it as starting point in CG in the current iteration. However Martens (2010) only tried it with Gauss-Newton approximation to Hessian we decided to also give it a try with Fisher Information Matrix.

Using full step direction from previous iteration in initialization of conjugate grad led to slightly better convergence rate and visible variance reduction of rewards.

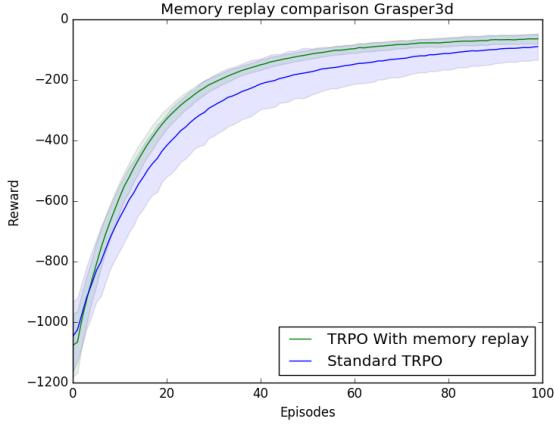


Figure 8: Memory replay convergence comparison on Grasper3 environment.

Parameter setting of experiments was exactly the same as in replay memory section only the algorithm changed a little.

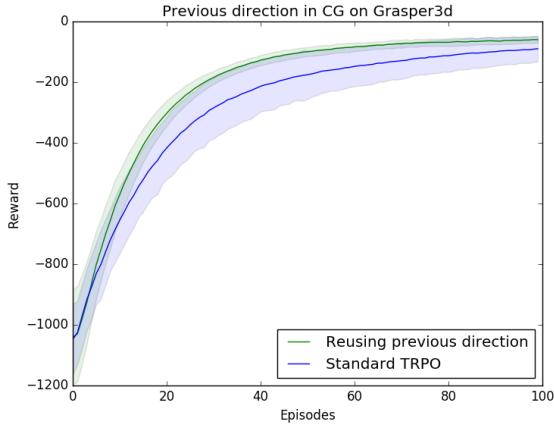


Figure 9: Reusing previous gradient direction convergence comparison on Grasper3 environment.

5.3 Combining both approaches

Assuming that both of the previous methods improves convergence rate we decided to combine them with the goal of maximizing usage of previously acquired information. Results shown in Figure 10 suggests that combining both approaches might not be a good idea and that both approaches do not play well together. It might be interesting to study why these two approaches do not work together and achieve better convergence property than either of them alone.

5.4 Parallelisation of TRPO

Nowadays when CPU frequency is no longer increasing parallelism is becoming hot topic in computing. There-

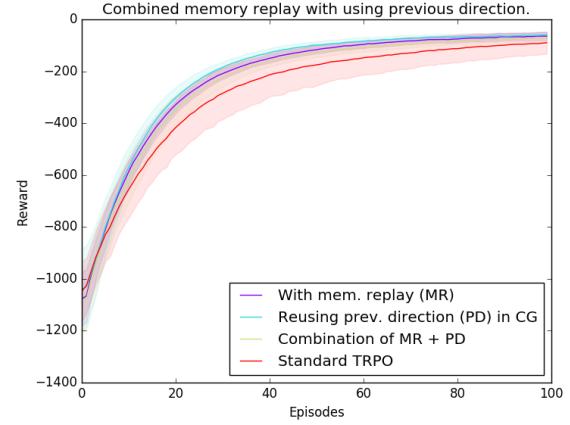


Figure 10: Combination of both convergence improving techniques on Grasper3 environment.

fore using multiple processors for our training was of the shelf method for increasing computation speed. We have observed that we spend more than 95% of our computational time sampling paths and simulating rather than computing gradient and updating network. Gathering multiple trajectories on different cores is inherently parallel task. That means there is no computational dependency of one from other. Knowing this we expected the speed up to be almost linear.

Using this knowledge we implemented parallel procedure for sampling trajectories by instantiating multiple simulation environments and multiple copies of policy network. In Figure 11 we depict the speedup of two processors compared to a single processor. In Figure 12 we compare one with five parallel processes.

Figure 11 was generated on computer with 2 core Intel Core i5 with 2,4GHz while for Figure 12 we used 6 core AMD Phenom II X6 1090T with 3.2GHz. We

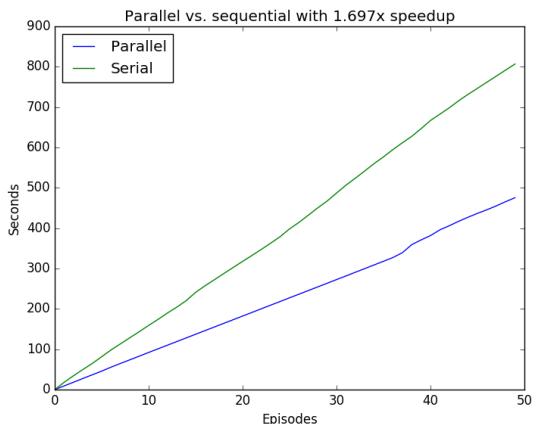


Figure 11: Sequential mode vs parallel mode with two processors. It was generated on Reacher2d environment with 300 rollouts per episode

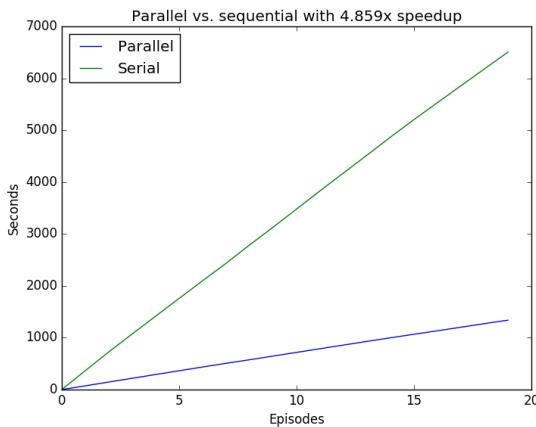


Figure 12: Sequential mode vs parallel mode with five processors. Generated with Grasper3D environment with 1000 rollouts for each episode

have not achieved expected linear speed up which could be due to multiple reasons. In the one vs. two processors we trained on personal laptop which processors were used by several other applications. This could take some of its computational power. Another part of computational time was taken by merging trajectories from different processes together. Second experiment with 5 cores seems much better. It is probably caused by the fact that we have 6 core processor and the last core can be used for computations required by other processes.

6 Conclusion

We have demonstrated that it is possible to learn robotic reaching and grasping by applying pure forces/torques on robots DoFs with reinforcement learning. We have also described how to solve reaching behind wall. Then we proposed convergence improvements on standard TRPO algorithm by using information from previous iterations and showed how speed up TRPO computational time with parallelism.

Acknowledgment

This research was supported by Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava by purchasing the license of an efficient MuJoCo simulator, used in our experiments. We were also supported by KEGA project no. 017UK-4/2016.

References

- Amari, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276.
 Bohg, J., Morales, A., Asfour, T., Kräigc, D. (2014).

Data-driven grasp synthesis - A survey. *IEEE Transactions on Robotics*, 30(2):289–309.

Castellini, C., Orabona, F., Metta, G., Sandini, G. (2007). Internal models of reaching and grasping. *Advanced Robotics*, 21(13):1545–1564.

Glorot, X., Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS’10)*.

Kalakrishnan, R. et. al (2011). Learning force control policies for compliant manipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Kraft, D., Detry, R., Pugeault, N., Baseski, E., Guerin, F., Piater, J., Krüger, N. (2010). Development of object and grasping knowledge by robot exploration. *IEEE Transactions on Autonomous Mental Development*, 2(4):368–383.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D. (2015). Continuous control with deep reinforcement learning. *Computing Research Repository*, abs/1509.02971.

Lin, L.-J. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8(3):293–321.

Martens, J. (2010). Deep learning via hessian-free optimization. In *Proceedings of the 27th International Conference on Machine Learning*, 735–742.

Pascanu, R., Bengio, Y. (2014). Revisiting natural gradient for deep networks. In *International Conference on Learning Representations*.

Pearlmutter, B. A. (1994). Fast exact multiplication by the hessian. *Neural Comput.*, 6(1):147–160.

Schulman, Moritz, e. a. (2015). Trust region policy optimization. In *International Conference on Machine Learning (ICML)*.

Schulman, J., Moritz, P., Levine, S., Jordan, M. I., Abbeel, P. (2016). High-dimensional continuous control using generalized advantage estimation. In *International Conference of Learning Representations (ICLR)*.

Sutton, R. S., Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

van Hasselt, H., Wiering, M. A. (2009). Using continuous action spaces to solve discrete problems. *International Joint Conference on Neural Networks*, 00:1149–1156.

Zdechovan, L. (2012). Modelovanie uchopovania objektov pomocou neurónových sietí v robotickom simulátore iCub. Masters thesis, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava.

Metafora a literárna recepcia

Jana Kuzmíková

Ústav slovenskej literatúry SAV
Konventná 13, 811 03 Bratislava
jana.kuzmikova@gmail.com

Abstrakt

Cieľom príspevku je opísať metaforu v literatúre ako blend, zmiešanie, nový mentálny priestor vytvorený na základe zdroja a cieľa. Vytváranie blendov v individuálnej mysli čitateľa má výrazný dosah na konštruovanie textových významov a celkovú recepciu literárneho textu. Teoretické modely (Fauconnier, Lakoff) možno dobre ilustrovať na textoch modernistického prozaika Ivana Horvátha. Ilustračné príklady zároveň vedú k nastoleniu hypotézy o tvorivom prepojení metaforických procesov, čitateľskej empatie a recepčného spracovania textu (napr. ako čitateľský zážitok, „dojem“) aj s uvedomením si významov príbehu alebo lyrickej výpovede v zmysle osobného pochopenia diela a poučenia sa z prečítaného.

1 Úvod

Mentálne priestory sú „interfejsom“, styčnou plochou v ľudskej mysli, cez ktorú komunikujú dva systémy, človek a svet. Obsahujú domény, budované v diskurze za účelom, aby poskytovali kognitívny podklad na porozumenie a usudzovanie.¹ Vhodným materiálom na demonštráciu vytvárania, funkcie a zmiešania nových mentálnych priestorov v myslení sú nepochybne literárne texty, najmä ak sú tvorené na asociatívnom princípe. Moderná avantgardná literatúra sa často spoliehalá na princíp asociatívnosti, rátajúc s vnímanosťou, vkladom čitateľa pri recepcii textu. Jedným z typických diel v tomto zmysle je novela *Laco a Bratislava* (1926) od modernistického novátora Ivana Horvátha. Próza hýri impresiami, buduje ju rozprávčov asociatívny prístup k svetu. Svedčí o tom aj údaj,

že v texte sa vyskytuje vyše 110 zmienok a citácií v škále od Descarta, Schopenhauera a Nietzscheho cez Štúra, Novomeského, Smreka, rôzne dejinné a politické osobnosti a svetových spisovateľov Danteho, Apollinaire, Conana Doyla a Oscara Wilda, maliarov Gauguina či Benku až po hudobných klasíkov i džezmenov. V protiklade k širokému kontextuálnemu autorskému gestu je príbeh protagonistu Laca modelovo jednoduchý: ide o prerozprávanie hrdinovej prvej vážnejšej lásky počas jeho vysokoškolských štúdií v Bratislave.

2 Konvenčná metafora

Pozoruhodný je hned' začiatok Horváthovej prózy *Laco a Bratislava*:

„*Krajina bola ako obraz*, vrchy sa roziahli, dostali ladné obrysy, hádam aby ich Benka mohol tak kresliť. Jeseň namaľovala hrubým štetcom červené a tmavohnedé škvurny na úbočiach, zelen hôr bola celkom čierna ako oči zlej vily. Ludia skrčení vyberali niečo zo zeme, boli skrčení, lebo ináč by nezapadli do **rámca obrazu**. (...) Bol to **obraz**, ktorý vždy Laca dojal. Stál na skale, tesne pri rieke, v krvi sa mu ozýval hlas predkov, ktorí za Rudolfa odsekávali hlavy Turkom, za čo dostali címer s hviezdom a polmesiacom. Jeho predkovia zemania strážili kedysi Váh, mali krivé meče a smelé srdcia. Laco nemal meč, ale mal srdce zrazu plné túžby za niečim neznámym, bol by tiež strážil Váh, ale nemal pred kým.“ (Horváth, 2010, 102, podč. J. K.)

Autor uvádzá svoje rozprávanie predstavou krajiny ako obrazu i výseku, pre ktorý je dôležitý rám. Hrdina Laco sa vnímaným obrazom dojima, cíti pred rodným krajom bázeň, ale zároveň ho túži opustiť, vyjst' za jeho hranice do neznáma: „*Laco cítil, že je už v moci cesty, že sa musí pohnúť z miesta, kde žil od narodenia*“ (103). Nová životná púť mladého hrdinu sa začína realizovať nedlho po vzniku prvej Československej republiky (1918), keď sa vyberie na vysokoškolské štúdiá do Bratislavu. Spoznáva „mestský“ svet, nasáva modernú domácu i svetovú kultúru a zažije prvú lásku.

¹ Mentálny priestor teda definujeme ako doménu v diskurze. Doména, napríklad *krajina*, je vnútorné štruktúrovaná rámcami, resp. nosníkmi (napr. prechádzat' sa v prírode alebo maľovať krajinu) a kognitívnymi modelmi (napr. ladné obrysy vrchov, na poli pracujúci človek, dojímavý obraz atď.). Idealizované kognitívne modely vznikajú z idealizovaných črt kategoriálnych objektov. Do jednej kategórie patria objekty, ktoré sú si blízke, podobné, previazané, uzavorené voči iným kategóriám a majú príbuznú funkciu. Na podstatných črtách kategoriálnych objektov sa budujú prototypy (bližšie Lakoff, 2006).

Sčasti autobiografický Horváthov príbeh „vystupuje“ z metaforicky „zarámovanej“ považskej krajiny, ktorá je rozprávačovi detsky blízka, vyvoláva v ňom dojatie. Pokračovaním, výstupom za „rám“ obrazu je metaforická schéma životnej cesty, spojenej so zvedavou túžbou po nových, neznámych svetoch. Podľa Fauconnierovej teórie mapovania vieme štruktúrne opísť obrazové schémy oboch domén, resp. ich rámce: ohraničenú krajinu i z nej vychádzajúcu cestu. Činným prvkom, rolu v oboch rámcach je mladý hrdina Laco, hoci v prípade krajiny je jej aktívnym agensom (podávateľom, konceptualizátorom obrazu) a v prípade cesty je pasívnym činiteľom (je v moci cesty):

Laco v rodisku	→	Laco na ceste
rodný kraj		cesta ----- >

Predstavy rodnej (známej) krajinu ako obrazu v ráme (obraz v ráme je príkladom elementárnej obrazovej schémy nádoby) a cesty (príklad elementárnej obrazovej schémy trajektórie) ako prechodu do nového sveta sú tzv. konvenčné metafory, ktorých tvorivý potenciál si dnes už ani neuvedomujeme, lebo čím je text prototypovejší, tým menšia je interpretačná aktivita čitateľa. To však neznamená, že nie sú v našom rozmysľaní a vyjadrovaní neustále prítomné, so živými inferenciami voči reálnemu svetu. Ako zdôrazňuje G. Fauconnier, nestaci sa zameriavať len na „autonómnu“ štruktúru jazykového prehovoru, rovnako dôležité sú prináležiace nelingvistické konštrukcie (Fauconnier, 2006, 4). Z hľadiska pochopenia významovej štruktúry, čiže vzhľadom na prebiehajúcu komunikáciu a diskurz, sú určujúce aj sociálne a kultúrne súvislosti. K nim sa vzťahujú i citové položky, integrované v doménach, ich rámcach a rolách: pri čítaní citovaného úryvku zaznamenávame nielen obrazový výrez z Považia, ale aj Lacovu *lásku* k rodnému kraju a jeho nová životná etapa je iniciovaná *pocitmi* túžby, zvedavosti, nutkavého vykročenia za neznámymi, tajomnými obzormi. V diskurze je teda dôležité najmä to, čo je takpovediac v „centre diania“. V priebehu mentálnych priestorov sú to tie oblasti, ktoré sú v ohnísku (focus) komplexných kognitívnych konfigurácií (kognitívnu konfiguráciu tvoria navzájom pospájané domény, resp. mentálne priestory). Štruktúru metaforických procesov priblížim na nasledovnej, tentoraz už originálnej Horváthovej metafore.

3 Inovatívna metafora

3.1 Zdroj, ciel, zmiešanie (blend)

Metaforické procesy možno dobre analyzovať na inovatívnej metafore:

„Z každého obchodu doliehalo na ulicu zvuky detského pláču a akordeónu, ulice boli od nich šmykľavé.“ (Horváth, 2010, 125)

Môžeme začať tradičným opisom, že v recepcii tejto metafory evidujeme **zdroj/e: detský pláč a akordeón a ciel: ulicu**. Ako však z detského pláču a zvukov akordeónu môžu byť ulice šmykľavé? Ako túto metaforu čitateľ pochopí?

Pláč, hudba a šmykľavosť, teda čiastkové vlastnosti a prvky metaforickej štruktúry svojím jednoduchým spojením, pridružením neurčujú, prečo by ulica mala byť kvôli zvukom pláču a akordeónu šmykľavá. Vysvetľuje to len ich projekcia (projection mappings) do emergentnej štruktúry zmiešania (blendu). Stane sa tak v mentálnom priestore zmiešania, ktorý vzniká čiastočným pospájaním, prekrytím východiskových domén z tzv. vkladov (inputs).

Štruktúra tejto metafory je nasledovná:

vklad 1 (**zdroj 1**): *detský pláč* - zvuk
 vklad 2 (**zdroj 2**): *akordeón* - zvuk
 vklad 3 (**ciel**): *ulica* – šmykľavá („hmat“)

Podľa vkladov vieme pri spolupôsobení tzv. background knowledge (predchádzajúce, pohotovostné poznanie v podobe rámcov, kognitívnych modelov, kultúrnych modelov, ľudových teórií atď.) určiť rámce (frames):

Rámeček (frame) vkladu 1: dieťa, pláč – zvuk, slzy
Rámeček (frame) vkladu 2: muzikant, akordeón – zvuk, naťahovanie a stláčanie mechu
Rámeček (frame) vkladu 3: ulica, šmykľavá – vlnká

Z troch vkladov (inputs) sa v integrovanej štruktúre mentálneho priestoru zmiešania (blend) ocitajú čiastkové štruktúry zo vstupných domén:

Z prvého rámcu, ktorý patrí do **domény pláču** prechádza do blendu *tahavý zvuk pláču a slzy, ich mokrost*. Z druhého rámcu v obsahu **domény hry na hudobnom nástroji – akordeóne** patrí do štruktúry nového mentálneho priestoru *tahavý zvuk harmoniky* a jeho produkcia – *kľzavý pohyb mechu*. Z tretieho rámcu v **doméne ulica** prechádza do blendu *šmykľavosť ulice*.

3.2 Základné formovanie blendu

Šmykľavosť ulice je v blende dostredivo súčasťou **emergentnej štruktúry**, ktorú zakladajú elementy:

1. pláč – tahavý pláč, slzy: atribút mokrost
2. hudba – kľzavý pohyb mechu aj muzikantovej ruky, tiahla melódia
3. ulica – šmykľavosť

vlnkost (1.) + *tahavý sklz* (1., 2.) <-> *šmykľavá ulica* (1., 2., 3.)

Šmykľavá ulica v synestetickom blende a za spolupôsobenia predchádzajúceho, záložného poznania znamená neistotu chôdze a smeru, vybočenie z nasmerovania. Šmykľavosť *späťne* evokuje nesplnené želania detí, ktoré azda nedostali v obchode to, čo chceli a preto pláču, i tiahlu, žalostnú hudbu harmonikára. Recepciu danej metafory ako koherentnej scény podporuje a usmerňuje vnútorná, mentálna reprezentácia minulej skúsenosti recipienta aj tak, že vo význame novej emergentnej štruktúry sú negatívne emócie neistoty, padania, zmaru.

4 Metaforické procesy ako kognitívne procesy

Ak aj pociťujeme pri inovatívnej metafore intenzitu výrazu v porovnaní s neuvedomovaným používaním lexikalizovaných („mŕtvych“) metafor, v princípe nejestvuje medzi štruktúrnymi procesmi týchto dvoch druhov metafor formálny rozdiel. Špeciálny, originálny prípad metafory vyplýva zo všeobecne aplikovaných kognitívnych principov a preto sú inovatívne metafory neraz rozvedením konvenčných, triviálnych metafor.² Napríklad originálna metafora, ktorú skrátene nazvime „šmykľavá ulica“, je podmienená lexikalizovanou metaforou *život je cesta*. V štruktúre metafory *život je cesta* je aj význam neistej, nestabilnej, kľukatej cesty (chôdze), napríklad prostredníctvom kvalitatívneho hodnotenia „často sa v živote pošmykol“. Paralelou i príkladom abstraktnejšej predstavy cesty, t. j. trajektórie (trajektória je elementárna obrazová schéma) je ulica. Metafora „šmykľavá ulica“ môže byť v prvej chvíli možno aj nezrozumiteľná, skôr akoby zameraná na efekt, ale jej prvotná nejasnosť či nespojitosť je postavená na ustálených štruktúrach konvenčnejších, konceptuálne a gramaticky zakorenencích metafor: *život je cesta, kľzavá melódia, tahaný pláč*. Fyzikálne atribúty zvuku *tiahlosť a kľzavosť* (vlnenie) a fyzikálny atribút ulice *vlhkosť* sa integrujú do jednej metaforickej predstavy, ktorá zdôraznené atribúty z troch vkladov obsahuje, ale ich aj presahuje vo významoch životnej neistoty, nenaplnených cieľov, žiaľu atď. Vďaka zbežným, už neuvedomovaným, ustáleným metaforickým procesom, mapovaniu medzi doménami (nielen medzi podobnými prvkami zdroja a cieľa!) a s prispeňím minulého, záložného poznania (background knowledge) dokážeme pochopíť i tvoriť nové inference, koncepty a emergentné štruktúry ako nástroje poznávania a jeho rozvíjania.³ Nové poznávanie sa na princípe zmiešavania (blending), založeného na schémach, rámcoch a modeloch, rodí neustále, synchrónne.

Je len ilúziou ľudových teórií, že jazykový výraz má význam, ktorý spoločne vnímame v zásade rovnakým spôsobom vo vzájomnom porozumení. Konštruovanie blendov a vytváranie konfigurácií má

svoje pevné pravidlá,⁴ ale význam nie je obsiahnutý v gramatických jazykových štruktúrach. V skutočnosti sa jazykové výrazy u jednotlivých užívateľov (národov, kultúr) zväzujú s rozličnými a rozdielnymi kognitívnymi konfiguráciami. Platí to nielen pri metaforách, ale všeobecne v komunikácii. Význam je na primárnom základe podporovaný a vedený nielen špecificky jazykovými, ale všeobecnými kognitívnymi operáciami. Príčinou je fakt, že zmiešané mentálne priestory, ktoré neustále tvoríme v kognitívnych procesoch, nie sú detailné, dokonale špecifikované, ale práve naopak, sú nedourčené a flexibilné. Z veľkej časti si ich neuvedomujeme, do vedomia sa dostanú len slová, emergentné významy a s nimi súvisiace pocity.

Súbor možných mapovaní pri recepcii metafory teda nie je len funkcioňa štruktúrnych a gramatických väzieb medzi exponovanými výrazmi. Pridanou hodnotou je, že konštruovaná štruktúra blendu je dynamická, emergentná a podložená rôznou minulou skúsenosťou u rôznych ľudí. Kontext diskurzu zahŕňa aj sociálny rozmer, pragmatické podmienky a aktuálnu situáciu, udalosti v reálnom svete. Čitatelia súce vytvoria azda podobnú zmiešanú štruktúru metafory „šmykľavá ulica“, ale každý pritom použije svoju skúsenosť v istej situácii, takže konkrétny mentálny priestor danej metafory je v konečnom dôsledku naplnený odlišnými dátami a predstavami. Zároveň však platí, že do zmiešaného mentálneho priestoru sa dostane len to, čo doň patrí. Toto aranžovanie metaforicky vystihuje veta z citovaného úvodu Horváthovej novely: „*Ludia skrčení vyberali niečo zo zeme, boli skrčení, lebo ináč by nezapadli do rámcu obrazu*“ (podč. J. K.).

5 Kognitívna konfigurácia

Aká je teda kognitívna kompozícia metafory „šmykľavá ulica“?

1. Mapovanie domén zo zdrojov na cieľ je štruktúrnou projekciou na princípe analógie. Analógia mapuje čiastkové štruktúry zdrojovej domény na čiastkové štruktúry cieľovej domény:

zdrojová doména plácu + zdrojová doména hry na akordeóne → doména šmykľavá ulica

Následkom čiastočného mapovania náprotivkov medzi vkladmi (2 zdroje, 1 cieľ) vznikajú prepojenia:

<i>tahavosť</i> (plaču)	<->	<i>kľzavosť</i> (melódie)	<->
<i>šmykľavosť</i> (ulice)			
		<i>kľzavý pohyb</i> (mechu a ruky)	<->
<i>šmykľavosť</i> (ulice)			
<i>mokrosť</i> (sľz)	<->		
<i>vlhkosť</i> (ulice)			

² Pozri Gáliková, 2014, 40.

³ Tamže, 32.

⁴ Pozri Fauconnier, 2006, 186.

2. Štruktúry z domén (vkladov) sú navzájom mapovateľné vďaka tomu, že sú príkladmi všeobecnejšej, **abstraktnejšej schémy**. Schéma je rámcem (frame) s rolami, ktoré sú napĺňané prvkami z jednotlivých domén:

Rámec (frame) domény 1: dieťa plače a vydáva zvuky a slzy (3 roly: dieťa, zvuk, slzy)

Rámec (frame) domény 2: muzikant hrá na akordeón, ktorý vydáva zvuk v dôsledku muzikantovho naťahovania a stláčania mechu (4 roly: muzikant, akordeón, zvuk, mech)

Rámec (frame) domény 3: človek je na vlhkej ulici vystavený šmyku (3 roly: človek, ulica, šmyk)

Rámce a roly navzájom vytvárajú bohatú inferenčnú štruktúru, vznikajúcú a dynamizovanú vplyvom minulej, záložnej skúsenosti recipienta, čitateľa. Explicitnú inferenčnosť zakladá *šmykľavosť*, na ktorú sa viaže *ťahavý pláč, slzy, klzavá hudba a sklz muzikantovho pohybu*.

3. Induktívne prepojenie domén prostredníctvom čiastkového mapovania:

vlhkosť (1.) + *ťahavý sklz* (1., 2.) <-> *šmykľavá ulica* (3., 2., 1.)

Špecifické čiastkové štruktúry z jednotlivých domén navzájom splývajú a korešpondujú v **abstraktnejšej schéme**:

Naplnenie ulice (ako nádoby) vlhkým a klzkým materiálom spôsobuje šmykľavosť ulice (ako cesty, trajektórie).

Na základe zmapovanej (extrahovanej) abstraktnej schémy môže dôjsť k rozšíreniu mapovania metafory flexibilným pripájaním ďalších domén. Napríklad zvuky akordeónu evokujú nielen ťahavú, ale i smutnú hudbu v "uplakanej" ulici; šmykľavosť ulice ako cesty evokuje pády človeka na ceste životom, a možno aj naschvál padajúceho smutno-smiešneho klauna atď. Uprednostňovanie niektorých mapovaní inými, čiže toho, čo sa môže dostať do „nádoby“, súvisí s kontextami a všeobecnou heuristikou.

Kognitívne konfigurácie si môžeme predstaviť ako siete či mriežky, pričom je významné, že v danom okamihu je pod tlakom kontextu a gramatiky zasiahnutý, produkovaný len jeden mentálny priestor (Fauconnier, 2006, 111). Je však tiež dôležité, že konfigurácie „dokážu“ rozčleňovať informácie na čiastky, ktoré sa môžu vzťahovať k rôznym doménam. Pre priebeh diskurzu (rozvíjanie kognitívnych konfigurácií a inferencií) je podstatné vedieť, ktorý mentálny priestor je bázou, východiskom, ktorý je hľadiskom, stanoviskom a ktorý je fokusom, priestorom, v ktorom aktuálne prebieha vnútorná štrukturácia významu. Pri spracovaní metafory „šmykľavá ulica“

- v prvej etape recepcie je bázou + hľadiskom „z každého obchodu doliehalo na ulicu zvuky detského pláču a akordeónu“ a fokusom „ulice bola od nich šmykľavá“;

- v druhej etape mapovania: prvotný fokus sa stáva hľadiskom „ulice bola od nich šmykľavá“ a projektuje

sa nový fokus s emergentnou štruktúrou „*šmykľavá ulica <-> vlhkosť + ťahavý sklz*“

- v tretej etape mapovania sa pomocou vytvorenej kognitívnej konfigurácie s fokusom s novou štruktúrou emeruje význam inovatívnej metafory.

Mobilizovaný fokus je zároveň blendom. Blend, zmiešanie sa rozvíja nasledovne (Fauconnier, 149 – 150): Po mapovaní naprieč mentálnymi priestormi a po vytvorení generického priestoru (ulica obsahuje ťahavý sklz a vlhkosť) vzniká blend a nová emergentná štruktúra, keď sa z generického priestoru aj troch vkladov čiastočne projektujú ich domény do priestoru zmiešania. Vzniká kompozitná štruktúra, ktorú v plnosti neobsahoval ani jeden z vkladov a zároveň sú v nej aj úplne nové časti:

a/ rola ulica obsahuje ťahavosť, vlhkosť, šmykľavosť a emergentnú vzťahovú, kompletizovanú schému *Naplnenie ulice (ako nádoby) vlhkým a klzkým materiálom spôsobuje šmykľavosť ulice (ako cesty)*

b/ pridáva sa nový, prerábajúci prvok *pošmyknutie (pád)* ako emergentné rozvinutie klzavého *pohybu* (mechu a ruky harmonikára)

c/ pridávajú sa ďalšie nové, upravujúce časti, ktoré sa pripájajú na abstraktnú schému zo záložného a pohovostného minulého poznania v procese prepracovania štruktúry už podľa jej vlastnej logiky (tragické i komické pády na ceste životom, príslušné emócie atď.).

Pri metaforických procesoch nie je ústredné primárne priradenie jednotlivých mentálnych priestorov, vkladov, resp. zdroja a cieľa (plač, hudba, šmykľavosť), ale pre porozumenie je dôležité rozčlenenie informácií a ich rozličné posúvanie podľa istého kódu, ktorý obsahuje bázu, hľadisko aj fokus. Množina čítaní metafory je daná množstvom rozličných prepojení v konfiguráciách mentálnych priestorov (projekčné operácie nie sú len jednosmerné, ale aj *spätné*) a tiež počtom mentálnych priestorov, možných a dostupných v príslušných kontextoch. Významová škála čítaní jednej vety alebo jednej metafory neodráža štrukturálnu alebo logickú viacnosť či viacvýznamosť danej jazykovej formy, ale odráža jej potenciál na budovanie mentálnych priestorov. Keby sme však nevedeli strategicky prepájať relevantné domény, prepracovať sa k nejakému metaforickému významu, vynaložená myšlienková procedúra by bola bezcenná. Stratégie na výber z alternatívnych mentálnych priestorov zahŕňajú gramatické, logické a pragmatické faktory, ale v zásade nie je jasné, ako sú stratégie nastavované. Zo skúseností môžeme nastoliť hypotézu, že istú úlohu tu zohrávajú emócie.

6 Metafora a emócie

Inovatívne metafory sú poväčšine pragmatickou anomáliou. Ich nezvyčajnosť sa prejavuje aj v recepcívnych pocitoch, obsiahnutých v mapovaných doménach,

ich rámcoch či nosníkoch (frames). Čím odťažitejšie a zdanlivo neanalogické domény sa mapujú, tým špecifickejšie (nekonvergentné, zmiešané) pocity môžu vzniknúť u čitateľa. V mentálnych priestoroch aj emócie, spolu s procesmi mapovania, prúdia rôznymi smermi. Môžu podnecovať alebo hatiť kognitívne operácie a koherenciu kognitívnych konfigurácií (pripomeňme, že predchádzajúce, záložné poznanie sprostredkúva inferencie aj vo vzťahu k reálnemu svetu). Cítením v procese čítania som sa zaoberala v článku *Porozumenie literatúre* (Kuzmíková, 2016). Demonštrovala som, ako v recepcii prechádza čítanie napriek štrukturálnym textovým aj pojmovým doménami. Vďaka týmto procesom získavajú pôvodne zaznamenané významy a schémy latentne tušené, resp. komplementárne nové citové zafarbenia. V tomto príspevku ma zaujíma špecifickejší problém: ako súvisí vynaložená empatia⁵ čitateľa s kognitívnymi procesmi a celkovým recepcným spracovaním textu?

Upozornila som už viackrát na emócie, pridružené k schémam v jednotlivých doménach rozoberanej metafory. Pocitové reakcie čitateľa na metaforu

,,Z každého obchodu doliehalo na ulicu zvuky detského pláču a akordeónu, ulice boli od nich šmykľavé.“

môžu byť rôzne. Môžu byť negatívne (smútok u emócie empatickejho čitateľa), tragikomicke, humorné (pre diskurzívneho čitateľa, ktorý si uvedomuje hrdinovu neskúsenosť až naivitu) alebo zasunuté, nerelevantné (literárny vedec, ktorý uvedenú metaforu analyzuje štrukturálnym spôsobom, si pridružené emócie ani nemusí všimnúť).

To, čo v recepcii zohráva kľúčovú úlohu, je mentálne fokusovanie, "centrum diania". Pri spracovaní metafory „šmykľavá ulica“ sa projektuje fokus s už známou emergentnou schémou:

Naplnenie ulice (ako nádoby) vlhkým a klzkom materiálom spôsobuje šmykľavosť ulice (ako cesty).
Pocity, pridružené k tejto schéme, sú funkciou priebežného (predchádzajúceho) i ďalšieho fokusovania kognitívnych procesov, zamerania čitateľa. Svoju úlohu v inferenčných procesoch zohráva čitateľský typ, ale tiež druh a sila investovanej empatie. Čitateľský typ a prejavnený druh empatie sú navzájom podmienené, ale nie úplne.

6.1 Emočná a kognitívna empatia

Všeobecne rozlišujeme dva čitateľské typy, resp. spôsoby čítania: sentimentálne alebo diskurzívne čítanie. S týmto delením súvisí tiež druh a miera investovanej empatie. Čitateľská empatia býva viac

⁵ Empatia je spontánnym zdieľaním pocitov, ktoré môže byť vyvolané, keď je človek svedkom emocionálneho stavu iného človeka, keď sa dozvie o rozpoložení niekoho iného alebo aj keď číta, resp. vníma umenie.

emočná alebo viac kognitívna.⁶ O kognitívnej empatii hovoríme vtedy, keď človek vie určiť alebo prisúdiť istý mentálny stav sebe či iným ľuďom, ale nevyhnutne ho nemusí zdieľať, preciťovať. Naopak, pri emočnej empatii sa jedinec citovo stotožňuje s iným človekom, literárnym hrdinom, spolucíti s ním. Emočná empatia podporuje sentimentálne čítanie literatúry, zatiaľ čo kognitívna empatia väčšinu súvisí s diskurzívnym, resp. pozorným čítaním.

Pri vnímaní metafory "šmykľavá ulica" sa čitateľ s rozprávačom Lacom bud' empatizuje alebo nie. Ak prežíva Lacove pocity (emočná empatia), potom sa automaticky nalaďí aj na melancholické evokácie citovanej metafory a pridružené negatívne významy pádu, životných nezdarov. Ak iný čitateľ vníma Lacov príbeh diskurzívnejšie (kognitívna empatia), dokáže sice pochopiť základnú emóciu smútku v "šmykľavej metafore", ale nepotrebuje (hoci môže) ju s Lacom prežívať. Prvotné negatívne pocity sa napríklad môžu komplementárne rovinúť až do tragikomickej scény šmykľavej ulice so smiešne padajúcimi ľuďmi (ako súčasť tragikomickeho pohľadu na život). Kognitívna empatia umožňuje aj zlomyseľné pocity, dokonca aj radosť z negatívnych emócií, sprostredkovaných rozprávačom Lacom.⁷

Ďalšou a emočne najchladnejšou možnosťou čítania "šmykľavej metafory" je jej doslovne významové čítanie, pri ktorom čitateľ nevie spontánne prepojiť metaforické domény, nie je schopný nájsť kód metafory, ale vníma skôr len jednotlivé slová- kategórie. Citové položky tak ostávajú zasunuté, hoci môžu v istej miere pôsobiť. Ne-empatické, odosobnené čítanie sa vyskytuje aj pri abstraktej štrukturalistickej analýze metafory. Ne-empatické čítanie však nemožno jednoducho označiť za ne-emočné, pretože môže byť nesené istými pocitmi (túžba odhaliť mentálne princípy fungovania metafory a pod.).

Posledným špecifickým príkladom mentálneho a citového fokusovania pri čítaní je, keď čitateľ sice číta intersubjektívne, ale je zameraný na seba. Nesúčíť s Lacom, ale smúti sám nad sebou, nejakou svojou skúsenosťou. Taktôto sa cez empatiu prenáša literárna fikcia do reálneho života. Spúšťa sa katarzia, čitateľ sa môže oslobodiť od vlastných negatívnych emócií.

7 Záver

Rozbor jednej metafory z novely Ivana Horvátha *Laco a Bratislava* aj napriek jej minimálnemu kontextovaniu ukázal, ako rozdielne môžu na metaforu reagovať čitatelia, ktorí do komunikačného procesu vnášajú

⁶ Uvedené všeobecné delenie čitateľských typov a čitateľskej empatie možno ďalej špecifikovať podľa širších psychologických, textových a kultúrnych kontextov. Pretože som sa v tomto príspevku sústredila na analýzu len jednej metafory "šmykľavá ulica", stačí nasledovne pracovať so základným delením čitateľských typov a empatie.

⁷ Bližšie Triebel, 2016, 128.

vlastnú osobnosť s istým poznaním a motiváciami v určitej situácii.

Vďaka tomu má hypotéza o tvorivom prepojení metaforických procesov, čitateľskej empatie a celkového recepčného spracovania textu reálny základ. Zážitok, „dojem“ z čítaného sa odvíja od toho, čo sa ocitá vo fokuse kognitívnych konfigurácií v mysli čitateľa počas recepcie - ale i po nej (celkové, konečné pochopenie umeleckej výpovede či poučenie sa z prečítaného). Pri oddychovom, resp. sentimentálnom čítaní možno viac predpokladať emočnú empatiu; pri diskurzívnom, resp. analytickom čítaní je prítomná skôr kognitívna empatia. V oboch prípadoch sa však na fokusovaní mentálnych priestorov podielajú aj pocity. V tomto zmysle majú aj pocity kognitívnu funkciu.

Poděkovanie

Štúdia vznikla v rámci grantového projektu VEGA 2/0078/14 Kognitívne orientovaná literárna veda. Teória a príkladové štúdie zo slovenskej literatúry. Riešiteľka Jana Kuzmíková.

Literatúra

Fauconnier, G. (2006). *Mappings in Thought and Language*. Cambridge University Press, 7. vyd.

Gáliková, S. (2014) Význam a funkcia konceptuálnej metafory z perspektívy kognitívnej vedy. V kolekt. monografii *Literatúra v kognitívnych súvislostiach*, str. 29-54.

Gärdenfors, P. (2014) *Geometry of Meaning: Semantics Based on Conceptual Spaces*. The MIT Press, 1. vyd.

Kuzmíková, J. (2016) Porozumenie literatúre. V zborníku *Kognice a umělý život XVI*, str. 101-104.

Kuzmíková, J. (2016) Čítanie, krása a čítanie. *Romboid*. LI (10): 111-115.

Lakoff, G. (2006) *Ženy, oheň a nebezpečné věci: Co kategorie vypovídají o naší mysli*. Triáda, 1. vyd.

Triebel, D. (2016). The complexities of literature, empathy, and prosocial behaviour. V kolekt. monografii *Schlüsselkonzepte und Anwendungen der Kognitiven Literaturwissenschaft*, str. 121-140.

A Model of the Reference Frame of the Ventriloquism Aftereffect using a priori bias

Peter Lokša, Norbert Kopčo

Institute of Computer Science, Faculty of Science, Šafárik University in Košice
peter.loksa@gmail.com, norbert.kopco@upjs.sk

Abstract

The reference frame (RF) used by audio-visual (AV) spatial representation is likely to be head-centered or eye-centered, aligned with the RFs of either the unimodal auditory (head-centered) or visual (eye-centered) representations. Results of previous RFAV studies are inconsistent, suggesting that the RF is either mostly head-centered, when examined in the periphery, or a mixture of head-centered and eye-centered, when examined in the central field (Kopčo et al., 2009; Lokša & Kopčo, 2016). Here, a model is proposed, assuming a form of a priori bias is combined with the adaptation due to AV stimuli. This model can explain the results in the baseline conditions, but not when ventriloquism aftereffect is induced. Therefore, additional mechanisms are likely to determine the AV RF.

1 Introduction

Vision plays an important role in how the brain processes auditory information (Alais & Burr, 2004). In the spatial domain, vision provides guiding signals for calibration of spatial auditory processing. This can be illustrated by the ventriloquism aftereffect illusion in which repeated pairings of spatially mismatched visual and auditory stimuli produce shifts in the perceived locations of sound sources that persist even when the sounds are presented by themselves (Alais & Burr, 2004; Knudsen & Knudsen, 1985; Knudsen & Knudsen, 1989). It might be that a supramodal spatial representation exists, directly used in motion planning etc.

The current study models data from a previous study which examined the RF of the ventriloquism aftereffect (RFVAE) (Kopčo et al., 2009). RFVAE might be identical or connected to RF of general supramodal spatial adaptation.

There were two basic hypotheses considering properties of RFVAE so would be: (1) head- and (2) eye-centered, in case of holding of which the RF is spatially fixated to specific body part (1) head itself (2) eyeball. The reason for choosing such ones as possible RFs is because respectively (1) auditory and (2) visual space is represented in these RFs (Brainard and Knudsen, 1995; Razavi et al., 2007).

In a previous study of the RFVAE, the observed aftereffect was compared between two conditions: eyes not shifted from the fixation point (FP) of ventriloquism aftereffect inducement, eye shifted to a new FP. By eye shift we mean change of fixation position, i.e., direction of the eye gaze when the stimulus is presented. It was hypothesized that if the aftereffect shifted with the eye shift, RFVAE would be probably eye-centered. If it didn't shift, it would be head centered, since head shifts neither. The goal of the previous modelling was to evaluate a possible mechanism causing this inconsistency of results with respect to the above hypotheses.

In the current study we first show the behavioral results, followed by an extension of the model and its evaluation.

2 Experimental data

The experimental data used here are taken from a previous study that investigated the reference frame of ventriloquism aftereffect (Kopčo et al., 2009).

2.1 Materials and methods

Fig. 1 illustrates the experimental setup and the hypothesized results. In the experiment the subject was sitting in a dark quiet room with his head fixed. The target speakers and LEDs (visual adaptor) were used to provide stimuli to the subject. The saccadic responses to stimuli were recorded.

To induce ventriloquism aftereffect the AV training trials with constant shift of light from sound were induced in specific azimuth region, while FPs of all such trials were same within session (training fixation point (TrFP); Fig. 1A).

To measure the aftereffect magnitude in the condition of eye not shifted from position of ventriloquism aftereffect inducement, the localization errors were identified according to responses to auditory-only (A-only) trials in TrFP in stimuli range -30° to 30°. Analogically was done for condition of eye shifted in so called Non-training fixation point (NTrFP). So within session AV trials were in TrFP and there were A-only trials in TrFP and A-only trials in NTrFP. These three kinds of trials were interleaved.

To see whether ventriloquism aftereffect is symmetrical or not, the session differed in (1) in shift of visual component of AV trials from its auditory component, and in training fixation point. There were three kinds of shifts of visual component: no shift (sound and light have same azimuth), positive shift (visual component is shifted by 5° to the side, toward which the TrFP is from 0°), or negative shift (opposite of positive shift). Regarding FPs azimuth axis can be flipped that TrFP would on 11.8° and NTrFP on -11.8° for each session.

Because discrimination abilities in center vs. periphery are inconsistent (Maier et. al., 2009), two different training regions of aftereffect inducement were used, but the same one within session. These two we call center and periphery. In Fig. 1 central one is shown.

The 9 speakers were displaced within same horizontal plane, while holding: distance of each speaker from center of the listener's head is equal; angle difference of the speaker from adjacent one is equal (7.5°) (see Fig. 1).

According to diff. in bias magnitude (bias of NTrFP A-only trials subtracted from bias of TrFP A-only trials; Fig. 1) the RFVAE had to be identified. Results are displayed in Fig. 2.

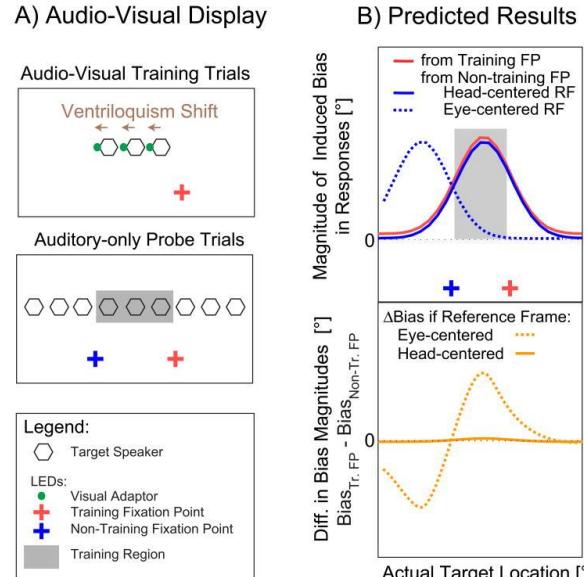


Fig. 1: A) Symbols in "Audio-Visual Training Trials" panel mark the azimuths of stimuli provided to subjects in audiovisual training trials, in the way that the azimuthal relative shift between physical location of stimuli, that are synchronous, are constant within given experimental session for each session. The symbols in "Auditory-only Probe Trials" panel mark azimuths of auditory-only trials, which were interleaved with the already mentioned training ones. B) This panel visualize hypothetical experimental data for cases of questioned reference frame being head- vs. eye centered. "Magnitude of Induced Bias in Responses"

here means localization error of them toward the shift in given session for each session.

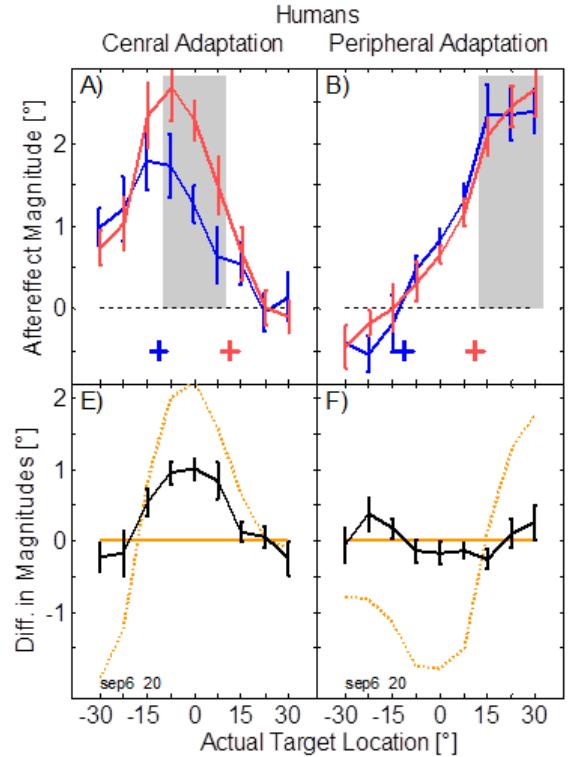


Fig. 2: Magnitude of ventriloquism aftereffect and reference frame determination according to difference between training vs. non-training trials. Red/Blue line - separation of probe auditory-only trials according to the pre-trial eye gaze azimuths (marked by '+' of given color). But eye gazes of all audio-visual training trials are preceded by the red one so this is called training fixation point (FP), and the blue one non-training one. The black line can be arithmetically described as the subtraction of blue line from the red line and we call it aftereffect FP dependence. The orange lines reflect hypothetical Aftereffect FP dependences: the solid one for the case of eye-centered head centered and the dotted one.

2.2 Data analysis

Fig. 2 shows that the result of RFVAE is inconsistent in the two regions: for central adaptation this RF seems to be mixed of head- and eye centered, while for peripheral adaptation it seems to be purely head-centered.

To resolve this inconsistency, we attempted to model the experimental data. To better understand the causes of the inconsistency, Fig. 3,4 show the detailed behavioral results for different conditions.

It is unlikely in the brain that two different forms of reference would be utilized for the same representation. On the other side, there are multiple other explanations for difference we observed, related to other forms of adaptation that might have occurred

in this experiment: the saccadic hypometria (undershoot), the expansion of auditory space, saccade adaptation.

The Kopčo et al. (2009) data showed another form of plasticity, described in the following section, for which the study has not been designed, and in modelling first two above-mentioned explanations were explored in previous article Lokša & Kopčo (2016).

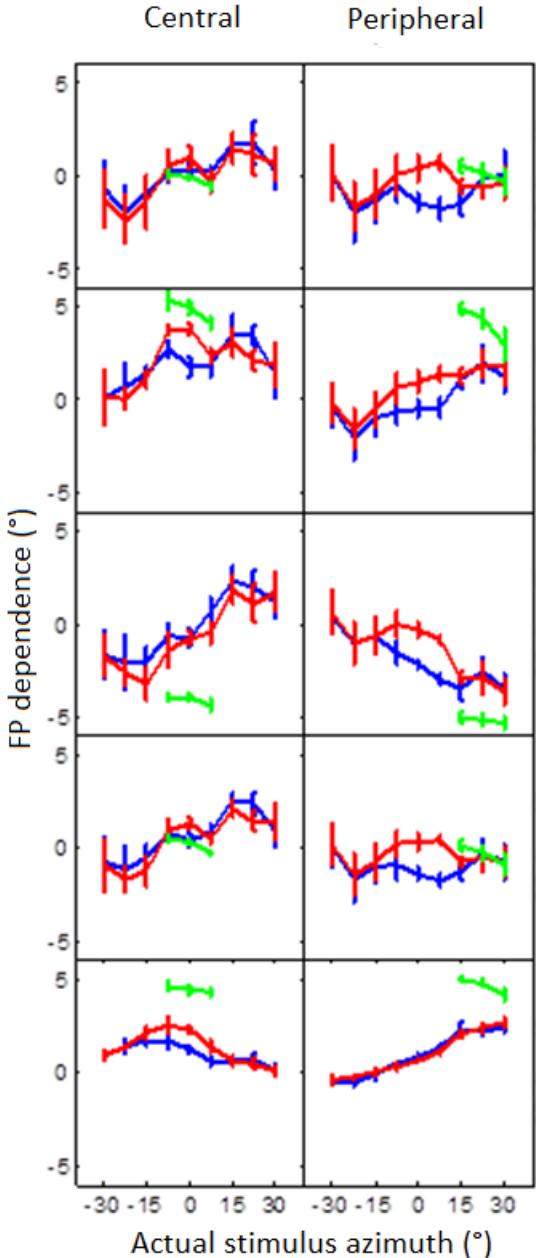


Fig. 3: Mean localization error of human subject experimental data and SEM across 7 subjects. Red line – A-only trials – training fixation point, blue line A-only data – non-training fixation point, green line – AV (training) trials, black line – difference between training vs. non-training A-only trial mean (FP dependence), magenta line – difference between peripheral vs. central adaptation FP dependence. Conditions according to rows respectively: 1. no shift,

2. positive shift, 3. negative shift, 4. mean across shifts, 5. aftereffect magnitude. The graphs in the 5th except of magenta lines row are little different with Fig. 2: A, B, E, F, except of yellow lines only because of technical errors and outliers removal.

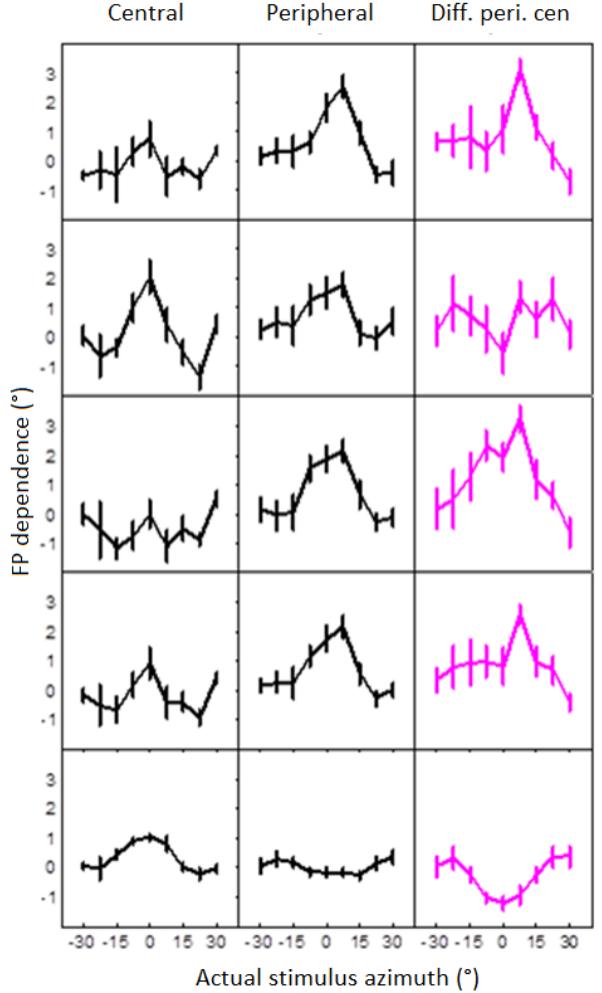


Fig. 4: Continuation of previous figure.

3 Unexpected form of plasticity

In Fig. 5 we observed inconsistency. In this figure we can see different azimuth and different condition that there are two types of cases for localization error being (1) depending (2) not depending on initial eye fixation point visualized as (1) similar or (2) dissimilar value of red vs. blue line: 1. all central azimuths, azimuths -30 to -15 in periphery and azimuths 15 to 30 in periphery. 2. azimuths -7.5 to 7.5 in periphery. This unexpected plasticity could be possible reason for inconsistency of central vs. peripheral RFs of ventriloquism aftereffect appearance.

In order to explain this inconsistency we attempted to model data present in this visualization (Typical property of this visualization is consistency of audiovisual training trials that affect localization errors

(so called no-shift) as the selection key for data included.

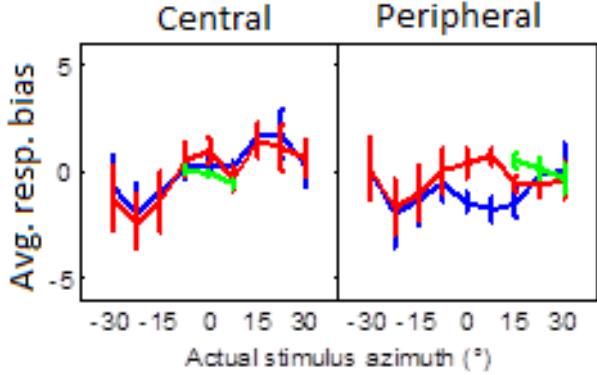


Fig. 5: Localization error for no-shift condition for different training regions (variables are in °).

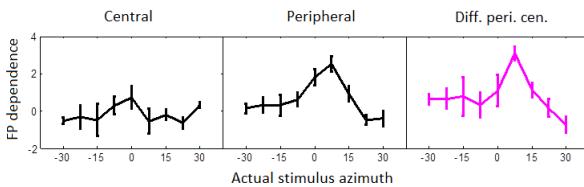


Fig. 6: Continuation of previous figure. FP dep. means difference between biases of TrFP vs. NtrFP. (The magenta graph for model of (Lokša & Kopčo, 2016) would be zero constant - black ones would be equal)

4 Modelling

This section presents a model of the newly observed adaptation, and also tests of relevant qualities of the model. The model assumes that a priori bias is modified by so called AV effect.

Previous modelling assumed that unexpected form of plasticity is caused by saccadic hypometry and expansion outside training region, specifically their additive composite, and this was proven to be insufficient to describe the data (Lokša & Kopčo, 2016).

4.1 Description of the current model

Basic idea of this modelling is a priori bias affected by biases of responses to AV stimuli (vertical position of green line). A priori bias is attracted by vertical position of green dots. Attraction is represented by weighted mean of A priori bias and vertical positon of green points. Weight of a priori bias is one of the free parameters (w_p). Weight of given vertical position of green point is product of (1) implicit free parameter ($1 - w_p$) and of (2) Gaussian function of: horizontal position of green point as the center of Gaussian function, azimuth of auditory-only stimulus as the main

input of it and another free parameter as the width of the Gaussian function (wdt).

A priori bias is sigmoid function modified to be odd (inflection point at vertical 0 instead of 0.5), where its horizontal center ($-FP \cdot c$), its height (hg) and its slope (sp) are adjustable by free parameters.

4.1.1 Established variables and functions

Free parameters (Tab. 1; parameters that were set by non-linear fitting algorithm of MATLAB (nlinfit)):

c	coefficient of horizontal position of inflection point according to fixation point for a priori bias.
hg, sp	height and slope of a priori bias.
w_p	weight for a priori bias.
w_{wdt}	width of AV effect.

Tab. 1: Free parameters.

Input variables (**Error! Reference source not found.**, vector of these variables determine prediction case):

a_{A-st}	actual azimuth of stimulus.
a_{FP}	azimuth of fixation point (red vs. blue line).
a_{AV-st}	vector of azimuths of A component of AV stimuli (these are marked as value of horizontal coordinate of green error bar centers).
$b_{AV-resp}$	vector of biases of response to AV stimuli (these are marked as value of vertical coordinate of green error bar centers).

Tab. 2: Input variables.

Established equations:

A priori bias is represented (left side of equation) and defined (right side of equation) in following equation (It is sigmoid moved vertically in order for have value domain in $[-hg, hg]$)

$$b_p(x, a_{FP}) = 2 \cdot hg \cdot [1 + e^{-z}]^{-1} - 1;$$

$$z = -sp \cdot [x - (-a_{FP} \cdot c)];$$

$$c, hg, sp > 0;$$

We mark the Gaussian function as following $Gauss(x, \mu, \sigma)$; where x is the main variable, μ is center and σ represents width of this function.

We mark and define the width function:

$$f_{wdt}(x, \mu, \sigma) = \frac{Gauss(x, \mu, \sigma)}{\sum_{i=-8}^8 Gauss(7.5 \cdot i, 0, \sigma)}.$$

This function is the function of closeness of x to μ .

The main function is implemented below, and its result is bias of response to A-only stimuli and the model predicts this bias:

$$b_{A-resp} = \frac{F_1 \cdot w_p + \sum_{i=1}^{count_{AV}} F_2(i) \cdot w_2(i)}{w_p + \sum_{i=1}^{count_{AV}} w_2(i)} \text{ where}$$

$$F_1 = b_p(a_{A-st}, a_{FP});$$

w_p is free parameter.

$F_2(i) = b_{AV-resp}(i)$ (This is the value of green error bar center);

$$w_2(i) = (1 - w_p) \cdot f_{wdt}(a_{A-st}, a_{AV-st}(i), w_{wdt});$$

$$count_{AV} = 3.$$

4.2 Performance

For spatially congruent AV stimuli the model looks like Fig. 7. It was fitted by 'nlinfit' Matlab nonlinear regression fitting function.

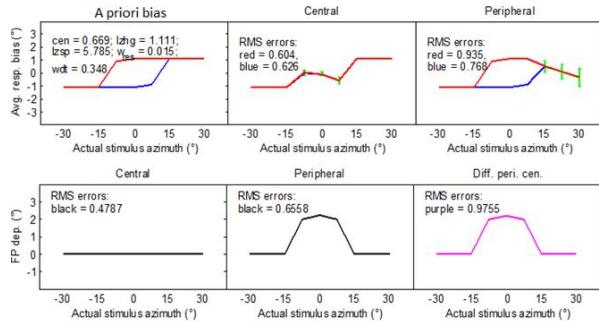


Fig. 7: Modelling results

Resulting coefficients for fitting on no-shift data:

$$c = 0.669,$$

$$hg = 1.111,$$

$$sp = 5.785,$$

$$w_p = 0.015,$$

$$w_{wdt} = 0.348;$$

Experimental data in Fig. 5 are well-explained by current model. You can see that for central adaptation the red and blue lines are almost equal and this is also

the case for prediction according to current model (Fig. 7). You can also see that difference in red vs. Blue line for peripheral adaptation is present in central 3 azimuths for both experimental data and prediction. Magenta line is also similar for experimental data and prediction according to model.

But the prediction of magenta line according to previous model (Lokša & Kopčo, 2016) would be zero constant for no-shift condition.

The model is less successful when we use fitting of also experimental data other than no-shift ones (Fig. 3 and 4). We did it and here are the results (Fig. 8 and 9).

The problem is that 5th row of these graphs, which displays difference of rows 2 and 3 divided by two, gives no difference between red and blue line, and that is inconsistent with experimental data (Fig. 3 and 4). This result shows the model's limitation in that the FP-specific shift (i.e., the difference between the red and blue lines in Fig. 8) is independent of the direction or magnitude of the visually-induced adaptation. Thus, the model cannot describe any adaptation that is eye-centered. This can be proven analytically, as shown in the following section.

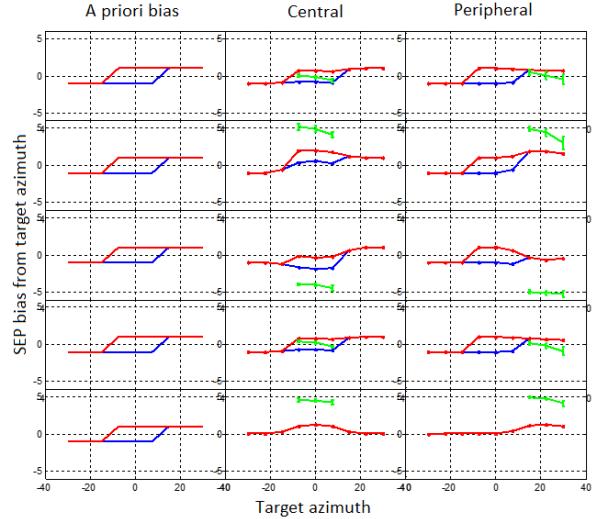


Fig. 8: Prediction of no-shift, positive shift, negative shift, mean across shifts, mean across shifts oriented as positive ones, respectively for given rows. Prediction was done on experimental data from Fig. 3 (Variables are in $^\circ$).

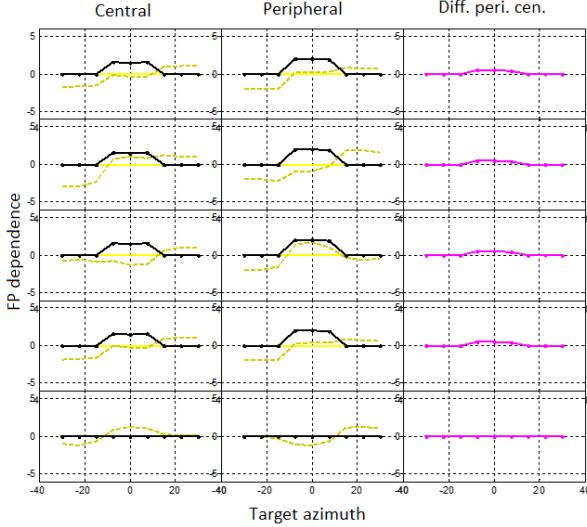


Fig. 9: Continuation of previous figure.

Resulting coefficients for fitting on data on all 3 shift conditions:

$$c = 8.980,$$

$$hg = 0.989,$$

$$sp = 8.930,$$

$$w_p = 0.260,$$

$$w_{wdt} = 4.939;$$

Mean square error of model fitted on all 3 shift conditions:

$$\text{MSE} = 1.375;$$

4.3 Proof of the inappropriateness of the current model

This section shows the weak aspect of current model in the manner that proves its weakness algebraically.

Model formula (bias to A-only response):

$$b_{A-\text{resp}}(a_{A-\text{stim}}, a_{FP}, a_{AV-\text{stim}}, b_{AV-\text{resp}}) =$$

$$= \frac{F_1 \cdot w_p + \sum_{i=1}^{\text{count}_{AV}} F_2(i) \cdot w_2(i)}{w_p + \sum_{i=1}^{\text{count}_{AV}} w_2(i)}$$

Main substitution I. (frequent variables merged):

$$o = \{a_{A-\text{st}}, a_{AV-\text{st}}\}$$

Main substitution II. (FP-dependence, difference of bias between different FP-s):

$$\begin{aligned} d_{FP}(a_{FP1}, a_{FP2}, & \text{bias}_{AV-\text{resp}}, o) = \\ & = b_{A-\text{resp}}(o.a_{A-\text{st}}, a_{FP1}, o.a_{AV-\text{st}}, b_{AV-\text{resp}}) - \\ & - b_{A-\text{resp}}(o.a_{A-\text{st}}, a_{FP2}, o.a_{AV-\text{st}}, b_{AV-\text{resp}}) \end{aligned}$$

Main hypothesis (FP-dependence independent of shift, thus also independent of $a_{AV-\text{resp}}$):

$$\begin{aligned} \forall(g, h, a_{FP1}, a_{FP2}, o) : \\ & : (d_{FP}(a_{FP1}, a_{FP2}, a_{AV-\text{resp}-g}, o) = \\ & = d_{FP}(a_{FP1}, a_{FP2}, a_{AV-\text{resp}-h}, o)) \end{aligned}$$

4.3.1 Inference

Substitution 1 (denominator):

$$B_d(o) = w_p + \sum_{i=1}^{\text{count}_{AV}} w_2(i, o.a_{A-\text{st}}, o.a_{AV-\text{st}}(i))$$

Substitution 2 (first member of numerator (a priori)):

$$B_p(a_{FP}, o) = F_1(o.a_{A-\text{st}}, a_{FP}) \cdot w_p$$

Substitution 3 (second member of numerator (AV effect)):

$$B_{AV}(b_{AV-\text{resp}}, o) = \sum_{i=1}^{\text{count}_{AV}} F_2(i, b_{AV-\text{resp}}(i)) \cdot w_2(i, \dots)$$

Inference 1: (Equivalent notation of model formula. To confirm, substitute members of current formula and compare)

$$\begin{aligned} b_{A-\text{resp}}(a_{A-\text{stim}}, a_{FP}, a_{AV-\text{stim}}, b_{AV-\text{resp}}) & = \\ & = \frac{B_p(a_{FP}, o) + B_{AV}(b_{AV-\text{resp}}, o)}{B_d(o)} \end{aligned}$$

Inferences 2 and 3 (Derived from main substitution II. and result of inference 1.) We see that d_{FP} is independent from $b_{AV-\text{resp}}$ variable in following two formulas (so independent of shift):

$$\begin{aligned} d_{FP}(a_{FP1}, a_{FP2}, & b_{AV-\text{resp}}, o) = \\ & = \frac{B_p(a_{FP1}, o) + B_{AV}(b_{AV-\text{resp}}, o)}{B_d(o)} - \\ & - \frac{B_p(a_{FP2}, o) + B_{AV}(b_{AV-\text{resp}}, o)}{B_d(o)} \end{aligned}$$

Thus:

$$d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp}, o) = \\ = \frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)};$$

Main hypothesis contradiction:

$$\exists(g, h, a_{FP1}, a_{FP2}, a_{A-stim}): \\ : [d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp-g}, o) \neq \\ \neq d_{FP}(a_{FP1}, a_{FP2}, b_{AV-resp-h}, o)]$$

Contradiction inference: (by substituting main hypothesis contradiction by result of inference 3)

$$\exists(g, h, a_{FP1}, a_{FP2}, azi_{stim}): \\ (\frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)} \neq \\ \neq \frac{B_p(a_{FP1}, o) - B_p(a_{FP2}, o)}{B_d(o)})$$

Contradiction of main hypothesis is disproved, thus the main hypothesis, that so called FP-dependence is independent of shift direction ($b_{AV-resp}$ depends on the shift direction), is proved. This fact can be visually seen on FIG. 9: 2nd, 3rd and 5th row.

5 Conclusion

We have described previous studies examining the reference frame of the ventriloquism aftereffect and its main results, which contain some ambiguity. We examined a part of the experimental data from that study, and we described a new adaptive phenomenon. We made the attempt to model these data and we have proven that the proposed model of a priori bias affected by AV responses seems appropriate to explain newly observed phenomenon when looking to the no-shift conditions, but inappropriate for the explanation of the difference of reference frames for shifted conditions (Fig. 2 and 4).

One of the alternatives for current modelling is use different weights for TrFP vs. NtrFP. Other one is to make NtrFP biases depending on TrFP biases instead directly of AV biases. Alternatively, completely other factors might play role. Additional modeling is currently required to examine these alternatives.

Acknowledgement

This contribution was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex

Environments and by the TECHNICOM project, ITMS: 26220220182, of the EU RDP.

Bibliography

- Alais, D. and Burr, D., (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol.* 14:257-262
- Knudsen, E. I. and Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science.* 230:545-548
- Knudsen, E. I. and Knudsen, P. F. (1989) Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience.* 9:3306-3313
- Kopčo, N., Lin, I., Shinn-Cunningham, B. G. and Groh, J. M., (2009). Reference Frame of the Ventriloquism Aftereffect. *Journal of Neuroscience.* 29(44): 13809-13814
- Lokša, P. and Kopčo, N., (2016). Modelling of the Reference Frame of the Ventriloquism Aftereffect. *Sborník z 16.ročníku konference Kognice a Umělý život (KUZ XVI).* 101-106
- Maier, J. K., McAlpine, D., Klump, G. M., Pressnitzer, D., (2009). Context Effect in the Discriminability of Spatial Cues. *JARO.*
- Razavi B, O'Neill WE, Paige GD. (2007) Auditory Spatial Perception Dynamically Realigns with Changing Eye Position. *Journal of Neuroscience.* 2007; 27:10249–10258.

Robot pred zrkadlom

Andrej Lúčny

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
Mlynská dolina, Bratislava
lucny@fmph.uniba.sk

Abstrakt

Prezentujeme riadiaci systém robota, ktorý interaguje so svojím obrazom v zrkadle za účelom pochopenia, že sa pozera sám na seba. Skúmame pritom z akých jednoduchších mechanizmov schopnosť uvidieť sa v zrkadle môže povstávať. Vychádza nám nasledovné. V prvom rade je to schopnosť vytvárať si model vlastného tela. V druhom rade schopnosť vytvárať analogický model z pozorovaného tela. Asociácia videného s vlastným telom potom môže vzniknúť na základe dokonalej korelácie medzi týmito modelmi. Potrebujeme však ešte nejaký mechanizmus, ktorý uvedie oba tieto modely do pohybu – dobre tu poslúži napríklad imitácia. Dôležitá je ďalej kategorizácia takýchto imitácií, typická pre členov tejakej society. Schopnosti rozlíšiť seba od ostatných tu podľa nás prechádza schopnosť rozlíšiť ostatných navzájom.

1 Úvod

Pri pohľade do zrkadla vieme, že sa pozéráme sami na seba. Zrkadlá nás obklopujú od malička, ale aj keď sa zo zrkadlom prvý krát stretne autochtoný obyvateľ pralesa, pomerne rýchlo zistí, že sa v zrkadle pozera na seba a to aj bez toho, že by mu niekto princíp zrkadla vysvetlil. Naopak deti mladšie než 18 mesiacov sa v zrkadle nevidia, ale napríklad hľadajú iné dieťa v priestore za zrkadlom (Obr. 1c) (Galup 1970).

1.1 Pohľad do zrkadla v prírode

Iba málo biologických druhov má rovnakú schopnosť. Nie je to vždy ľahké otestovať, avšak pomerne slušným dôkazom je úspech v tzv. dot-mirror teste, kedy sa najprv nechá testovaný subjekt so zrkadlom oboznámiť (aby vedel ako v zrkadle vyzerá) a potom sa skúma či na svojom tele hmatá po dodatočne (a nenápadne) urobenej bodke krikľavej farby na mieste, ktoré môže vidieť iba v zrkadle. Týmto testom úspešne prechádzajú všetky druhy ľudoopov, slony, delfíny, kosatky a straky. To nemusia byť všetky zvieratá, ktoré sa v zrkadle vidia, lebo napríklad krkavec je vo všeobecnosti považovaný za oveľa inteligentnejšieho ako straka, na rozdiel od nej však nemá záľubu v jagavých predmetoch, takže vykonáť s ním dot-mirror

test je ľažšie. Na druhej strane je možné, že u straky dot-mirror test dokazuje len to, že vie uchopovať jagavé predmety aj podľa obrazu v zrkadle a pritom nechápe, že sú umiestnené na jej tele. Dot-mirror test taktiež požaduje o niečo viac než samotné uvidenie sa (pod čím rozumieme pochopenie, že sa dívame na seba) a to vytvorenie zapamätanie si svojho obrazu.



a)



b)



c)

Obr. 1: Šimpanz, mačka a dieťa pred zrkadlom (šimpanz sa predvádzza, mačka útočne cerí zuby, dieťa mladšie ako 18 mesiacov hľadá iné dieťa za zrkadlom)

Takže, pokiaľ aj zviera neprejde dot-mirror testom, možno v určitých prípadoch oprávnene usúdiť, že sa v zrkadle vidí alebo nie, podľa pocitu bezpečia alebo naopak podľa pocitu ohrozenia, ktorý dáva najavo spôsobom charakteristickým pre svoj druh. Kým šimpanz sa zjavne pred zrkadlom predvádza (Obr. 1a), mačka má tendenciu na obraz v zrkadle zaútočiť (Obr. 1b).

Na základe poznatkov z prírody teda vieme, že schopnosť uvidieť sa v zrkadle je naviazaná na postupný vývin kognitívnych schopností jedinca a povstáva v určitom veku. Je v prírode zriedkavá a vyskytuje sa len u druhov, ktoré sa v istých ohľadoch ponášajú na človeka (ak vynecháme straky, tak všetky žijú sociálnym spôsobom života, majú rozvinutú schopnosť vnímania vlastného tela a majú veľký pomer medzi veľkosťou mozgu a tela).

1.2 Umelý pohľad do zrkadla

V tomto článku sa venujeme otázke z akých jednoduchších mechanizmov či modulov je možné schopnosť uvidieť sa v zrkadle poskladať. Nápovedou sú jednak uvedené poznatky z prírody, jednak svoje predstavy budeme testovať na umelom systéme, na ktorom sa budeme snažiť uvidenie sa v zrkadle demonštrovať.

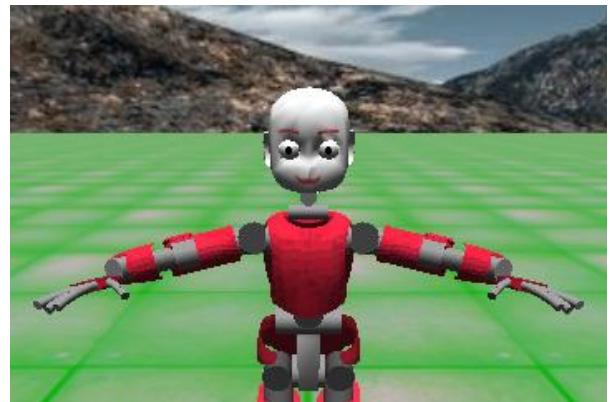
Zďaleka nie sme prví, kto sa o niečo také pokúšal. Vychádzame hlavne z práce Scasellatti a Hart 2012 (Scasellatti bol svojho času jeden z tvorcov robota COG v AiLab na MIT), podľa ktorých sa robot (alebo človek či zviera) v zrkadle vidí na základe dokonalej synchronizácie medzi pohybom tela a pohybom jeho obrazu v zrkadle. Ich prístup je pomerne náročný, lebo si dali za cieľ zostrojiť robota (Nico), ktorý dokáže konáť rôzne akcie podľa obrazu v zrkadle a to v priestore, ktorý inak nevidí.

Ich prístup považujeme za oveľa vhodnejší, než autorov robota QBO (Corpora 2011), ktorí vyhlasujú, že ich robot prechádza dot-mirror testom. Robí tak však na základe informácie od učiteľa, ktorý mu povie: "Teraz vidíš seba". Takže keby stretol svojho dvojníka, tak povie: "Toto som ja", zatial' čo Nico by vedel, že to nemôže byť on, lebo hoci vyzerá rovnako, hýbe sa ináč.

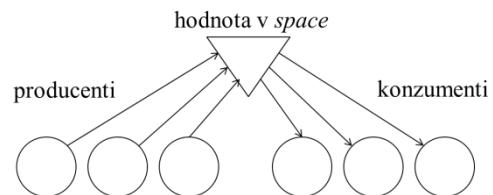
Na druhej strane z technického hľadiska by bolo pre nás príliš náročné urobiť robota ako je Nico a hlavne vyhodnocovať charakter pohybov tela pri autentickom modeli tela. Na skúmanie komponentov, ktoré za uvidením sa v zrkadle stojia, to našťastie nie je nevyhnutné, ako svojím prístupom ukázal Takeno 2008, pracujúci s veľmi jednoduchým robotom a jednoduchým modelom tela. My zvolíme strednú cestu, bližšiu skôr prístupu Takena: budeme pracovať s humanoidným robotom, avšak jeho pohyby výrazne obmedzíme, čím si zjednodušíme model jeho tela.

Konkrétnie budeme používať simulátor humanoidného robota iCub, zvaný iCubSim (Sandini a kol. 2007) (Obr. 2). Tento simulátor je pomerne kvalitný a poskytuje nielen ovládanie 53-och stupňov

voľnosti, ale aj ich monitorovanie, t.j. model tela. Je v ňom teda potenciálne rozdiel medzi pokynom, akú polohu má určitý klb zaujať a skutočnosťou, ktorá (ako-tak) rešpektuje zákony fyziky (používa fyzikálny engine ODE).



Obr. 2: Simulátor iCubSim



Obr. 3: Dátový tok v architektúre Agent-Space. (kruh označuje modul, trojuholník hodnotu v *space*)

1.3 Architektúra

Využívať budeme taktiež výsledky vlastnej predchádzajúcej práce v oblasti tvorby modulárneho softvéru, konkrétnie architektúru Agent-Space (Lúčny 2004). Táto nám umožní realizovať jednotlivé mechanizmy uvidenia sa ako na sebe nezávislé moduly pracujúce paralelne a vzájomne interagujúce prostredníctvom tzv. čiernej tabuľe (zvanej *space*). Tu nám dosť pomôže, že pri charaktere tejto architektúry zaniká potreba zaoberať sa faktom, že jednotlivé moduly majú rôznu rýchlosť spracovania údajov. Základnou myšlienkou toho, ako je to dosiahnuté, je, že kym jeden modul pravidelne prepočítava svoj výsledok a zapisuje ho na tabuľu, druhý ho odtiaľ pravidelne číta a nestaráme sa pritom, či nejaký výsledok prečítať nestihne. Podobne nevadí, že sa spracúvajú výsledky vypočítané viacerými modulmi v rôznom čase, hlavné je, že berie tie najnovšie. Jeden výsledok môžu pritom počítať viaceré moduly, pričom jeho odberatelia nevedia, od koho pochádza. No a pritom jeden producent výsledku môže mať (ale nemusí) vyššiu prioritu než druhý, takže výsledok od druhého sa berie v úvahu len vtedy, keď nie je k dispozícii výsledok od prvého. Každý výsledok môže mať (a spravidla aj má) ohrianičenú časovú platnosť, takže výsledok od prioritnejšieho producenta po čase zanikne a uvoľní

priestor pre výsledok od menej prioritného producenta, pokial ten prioritnejší nevyprodukuje výsledok ďalší (Obr. 3).

Táto architektúra je silne založená na myšlienkach subsumpčnej architektúry (Books 1999) a societného modelu myšle (Minsky 1986), avšak tieto myšlienky sú vyjadrené v jazyku multi-agentových systémov (Kelemen 2001).



Obr. 4: Simulátor iCubSim rozšírený o kameru

2 Implementačné prostriedky

Naším cieľom je poodhaliť podstatu procesu uvidenia sa v zrkadle pomocou implementovania umelého systému, v ktorom tento proces prebieha. Budujeme ho z modulov reprezentujúcich mechanizmy podielajúce sa na tomto procese, takže skúmame z akej štruktúry jednoduchších procesov tento proces povstáva.

V simulátore iCubSim simulovaný robot vidí virtuálnu scénu, ktorej je súčasťou, napríklad stôl s predmetmi. My ho integrujeme do systému, kde je rozšírený o fyzickú kameru, snímajúcu priestor pred monitorom, na ktorom je renderovaná postava iCub-u (Obr. 4). Vďaka tomu bude simulovaný iCub miesto svojej virtuálnej scény vidieť priestor pred monitorom, v ktorom sa nachádza programátor, t.j. človek. Pre naše účely je nevyhnutné, aby boli modely tela človeka a robota vzájomne porovnateľné – čiže nám vyhovuje, že iCub je humanoidný robot.

Ako sme však už spomenuli, model tela výrazne zjednodušíme, a to na jediné číslo – uhol náklonu hlavy na stranu (rovno 0° , doľava zhruba do -55° , doprava zhruba do $+55^\circ$). S ostatnými kĺbmi iCub-u nebudem hýbať a pohľad do virtuálnej scény s iCub-om zvolíme tak, aby na obrazovke monitora bola renderovaná len hlava.

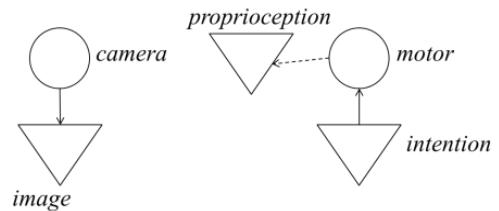
3 Mechanizmy

V zmysle zamýšľanej stratégie uvidenia sa v zrkadle, je našim cieľom implementácia systému, v ktorom by

bolo možné vyhodnotiť koreláciu medzi modelom tela, ktoré systém má a analogickým modelom tela, ktoré systém vidí. Myslíme pritom na biologickú relevantnosť komponentov, z ktorých systém skladáme, majúc na mysli špecifické vlastnosti tých biologických druhov, ktoré sa v zrkadle vidia. Systém budujeme inkrementálne, opierajúc sa o zvolenú modularitu. Pridaním ďalších modulov implementuje vždy ďalší mechanizmus, o ktorom predpokladáme, že za uvidením sa v zrkadle stojí.

3.1 Model tela (propriocepacia)

V prvom rade musíme mať k dispozícii obe porovnávané modely. Začneme s modelom vlastného tela robota. Ten už máme implementovaný v simulátore, takže ho musíme akurát prispôsobiť našej architektúre, t.j. dostať ho na onú čiernu tabuľu (do *space*). Implementujeme preto modul "motor", ktorý zo *space* číta želanú hodnotu nastavenia tela "intention", komunikáciou so simulátorom (prostredníctvom yarp rpc protokolu) sa ju snaží nastaviť telu robota, monitoruje skutočné nastavenie, ktoré sa podarilo dosiahnuť a zapisuje ho do *space* ako hodnotu "proprioception" (Obr. 5). Vďaka tomu vieme zápisom do *space* ovládať telo robota a čítaním zo *space* monitorovať skutočný stav jeho tela (Obr. 6).



Obr. 5: Implementačná fáza 1.



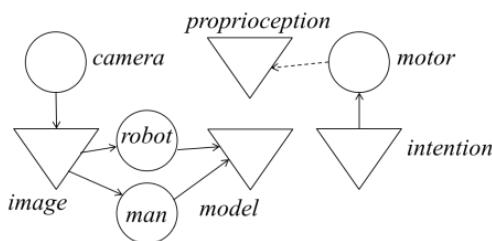
Obr. 6: Zjednodušený model tela robota (zobrazený ako vodorovná palička v čiernom poli dolu)

Paralelne k ovládaniu aktuátorov zabezpečujeme vstup zo senzorov. Náš jediný senzor je kamera, z ktorej obraz zapisujeme pravidelne do *space* (hodnota "image"). Moduly spracúvajúce tento obraz zdáleka nestihnú zo *space* prečítať každú snímku do *space* zapísanú, ale vďaka zvolenej architektúre sa tým nemusíme nijako špeciálne zaoberať.

3.2 Zrkadlenie

Teraz prejdime k získaniu analogického modelu tela, ktoré vidíme na obraze. Neriešime pritom, odkiaľ sa takáto netriviálna schopnosť (t.j. schopnosť popísat parameter svojho tela na podobnom tele videnom na obraze) berie, preste ju "nadrôtuje".

Kedže v našom experimentovaní zamýšľame, aby sa robot pozeral na obraz človeka i na svoj obraz (t.j. obraz robota), z oboch týchto obrazov musíme vedieť používaný model tela vytvoriť. Žiaľ, hoci iCub je humanoidný, jeho obraz sa od obrazu človeka natoľko lísi, že sa nám nepodarilo použiť na tento účel jednu spoločnú metódu. Avšak to nám v princípe neprekáža, lebo proces rozpoznávania si aj tak predstavujeme ako viacero nedokonalých metód zosúladených do funkčného celku. V našom prípade teda vyvinieme dve metódy, jedna bude spracúvať obraz robota, druhá človeka. Obe budú vytvárať rovnaký model tela. Avšak nielen to: ich výstupy budú zapísané na *space* do rovnakej hodnoty (hodnota "model" na Obr. 7). Keby sa teda ozvali obidve naraz (čo je málo pravdepodobné), navzájom by si svoj výsledok prepisovali. Každopádne však pritom ostatné moduly nemajú prístup k informácii od ktorej metódy hodnota modelu pochádza.



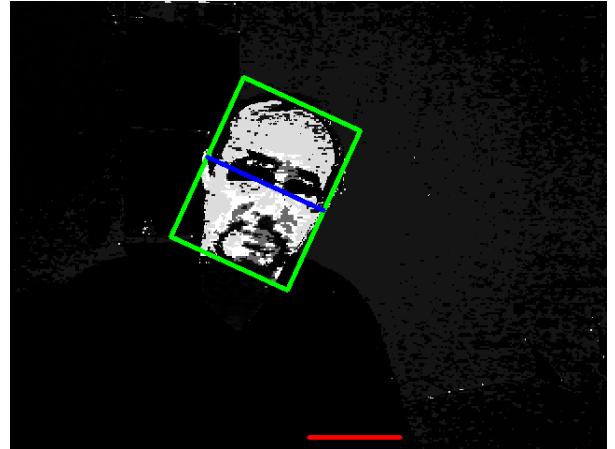
Obr. 7: Implementačná fáza 2



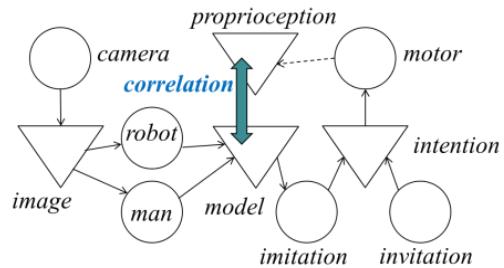
Obr. 8: Získanie modelu z obrazu robota (SURF)

Vytvorenie modelu z obrazu robota sme testovali tak, že sme pred kamerou otáčali vytlačený fotku monitora s renderovaným obrazom robota. Najlepšie sa nám osvedčilo použiť metódu SURF (Bay a kol. 2008). Tá na ľubovoľnom obraze nájde tzv. význačné body a opíše ich okolie tzv. deskriptorom a to nezávisle na posunutí, rotácii a zväčšení či zmenšení. To umožňuje

SURFu spojať význačné body s podobnými deskriptormi získané zo vzoru (obrázok tváre robota) a obrazu (aktuálny snímok z kamery). Z nich sa dá vypočítať projekcia vzoru na obraz a z tej už ľahko určíme uhol náklonu vzoru, čo je parameter zjednodušeného modelu tela, ktorý hľadáme (Obr. 8).



Obr. 9: Získanie modelu z obrazu človeka (Haar cascades a CamShift)



Obr. 10: Implementačná fáza 3

Kým jeden iCub je podobný druhému ako vajce, ľudia sú rôzni. Na rozpoznanie náklonu ich hlavy do strany preto použijeme inú metódu. Pokial je hlava vzpriamene, tvár dokážeme spoľahlivo identifikovať pomocou face detektora. Používame hotový detektor z OpenCV, založený na Haarových príznakoch v kaskádach (Viola a Jones 2001). Ide o klasifikátor naučený z pozitívnych a negatívnych prípadov. Jednotlivé príznaky si možno predstaviť ako binárnu masku malej veľkosti položenú na konkrétné miesto na obraz vzorov. Kaskády zase ako rozhodovací strom budovaný na základe čo najlepšej rozlišovacej schopnosti príznakov. Tento detektor tváre však funguje len vo vzpriamenej polohe. Pri náklone nám teda nepomôže. Poskytne nám však objekt na obraze, ktorý môžeme potom sledovať pomocou bežných metód (tracking). Používame metódu CamShift (Bradski 1998), ktorá je založená na histograme farieb sledovanej oblasti. Iteratívne sa potom vyhodnocujú body na aktuálnom obraze farebne najpodobnejšie iniciálnej oblasti. Sledovaná oblasť sa potom iteratívne posúva a rotuje podľa ďalšika oných dostatočne

podobných bodov na nových snímkach. Z rotácie sledovanej oblasti potom priamo dostávame potrebný model. (Obr. 9). Všetky použité algoritmy sú implementované v OpenCV (www.opencv.org) (Bradski 2000).

```
class imitationAgent : public Agent
{
private:
    string modelInclinationBlock;
    string inclinationBlock;
protected:
    void init (string args);
    void sense_select_act (int pid);
public:
    imitationAgent (string args) : Agent(args) {};
};

void imitationAgent::init (string args)
{
    istringstream in(args);
    in >> modelInclinationBlock;
    in >> inclinationBlock;
    timer_attach(1000,1000);
}

void imitationAgent::sense_select_act (int pid)
{
    double undefined = 360.0;
    double inclination =
        space_read(modelInclinationBlock, undefined);
    if (inclination != undefined) {
        space_write(inclinationBlock,inclination,1500);
    }
}
```

Obr. 11: Ukážka kódu modulu realizujúceho imitáciu. Druhý argument `space_read` je preddefinovaná hodnota pre prípad, že hodnota v `space` nie je definovaná. Tretí argument `space_write` je časová platnosť v ms. (Kód je zjednodušený.)

3.3 Imitácia

Disponujeme už modelom svojho tela (hodnota "intention") aj modelom videného tela (hodnota "model"). Koreláciu medzi týmto dvomi modelmi je však možné vyhodnotiť len pokiaľ sa tieto hodnoty v čase menia. Potrebujeme preto uviesť obe telá do pohybu. Jedným z mechanizmov ako to zariadiť je imitácia.

Imitáciu môžeme veľmi ľahko implementovať, nakoľko oba modely sú kompatibilné. Je preto možné priamo kopírovať hodnotu "model" do hodnoty "intention" (viď Obr. 10). Keď potom robot uvidí obraz človeka alebo robota, dá hlavu do rovnakej polohy ako majú oni. Tento modul (voláme ho "imitation") je tak jednoduchý, že jeho kód môžeme poľahky uviesť (Obr. 11). Dôležité je všimnúť si, že hodnota "model" má obmedzenú časovú platnosť. Takže keď človek či robot zmizne z obrazu, po vypršaní tejto platnosti, hodnota zo `space` zmizne a v dôsledku toho sa hlava robota vráti

do vzpriamenej polohy, keďže práve to je preddefinovaná hodnota pre prípad, že v `space` nie je "intention" definovaná.

Takže teraz keď sa človek posadí pred kameru, iCub imituje jeho pohyby hlavy (Obr. 12). Pokial chceme aby aj robot človeka nejako vyzýval k imitácii, môžeme doplniť modul "invitation", ktorý v prípade, že vidí človeka, ktorého model sa moc nemení, čas od času vygeneruje vhodný pohyb hlavy robota. To spraví tak, že náhodne vygeneruje uhol a zapíše ho do `space` – do hodnoty "intention" – pričom sa mu zíde vyššia prioritá zápisu, než má modul "imitation".

Nič menej sa ukázalo, že na predvedenie uvidenia sa v zrkadle tento modul nie je nevyhnutný. Je to vďaka tomu, že určovanie modelu videného tela má len určitú presnosť, takže jeho hodnoty prirodzene kolísajú aj keď sa videné telo vôbec nepohnie. Tieto drobné odchýlky stačia na to, aby vyvolali imitačný proces, keď sa robot díva na svoj obraz v zrkadle.

3.4 Sociálne modelovanie

Imitácia nám poskytuje dostatočné dátá na vyhodnocovanie korelácie medzi modelom tela a modelom videného (viď obojsmerná šípka na Obr. 10). Na to, aby sa tvor pozerajúci sa do zrkadla uvidel, musí nejako tieto hodnoty kategorizovať. Je pomerne ľahké si predstaviť, že pre tvora žijúceho v nejakej society má význam vyhodnocovať, ktorý z jedincov society je viac a ktorý menej ochotný imitovať, takže s obrazom jedinca spája aj informáciu o miere korelácie medzi uvažovanými modelmi. Primárne teda rozpoznáva členov svojej society. V zrkadle potom uvidí ako keby ďalšieho člena society, ktorý má nenormálne vysokú mieru onej korelácie, t.j. absolútну ochotu imitovať. A to je on sám.

Pre naše účely sme túto kategorizáciu neimplementovali, ale zamerali sme sa na to, či v zázname dát vieme rozlíšiť situáciu, kedy sa robot díva na človeka a kedy na svoj obraz v zrkadle.

4 Výsledky

Postupne sme vyvinuli systém, do ktorého sme vložili všetky moduly, o ktorých sa domnievame, že hrajú určitú rolu v procese uvidenia sa v zrkadle. Keď teraz imitujúceho človeka pred kamerou nahradíme zrkadlom, robot imituje vlastný obraz v zrkadle (Obr. 13). Z úvodu sa chvíľu nič nedeje, ale onedlho nepresnosť v určení modelu videného tela spôsobí, že v "mysli" robota sa obraz mierne pohnie a on sa to podujme imitovať. Následne sa rozbehne imitačný proces, však oveľa pomalšie, sústredenejšie a vo výraznejšej miere (náklon hlavy bude väčší než obyčajne), než je tomu pri interakcii s človekom.

Pritom matematicky rozlísiť dátá nie je jednoduchá záležitosť a s učebnicovou definíciou korelácie tu nevystačíme. Nič menej, pohľad na priebeh

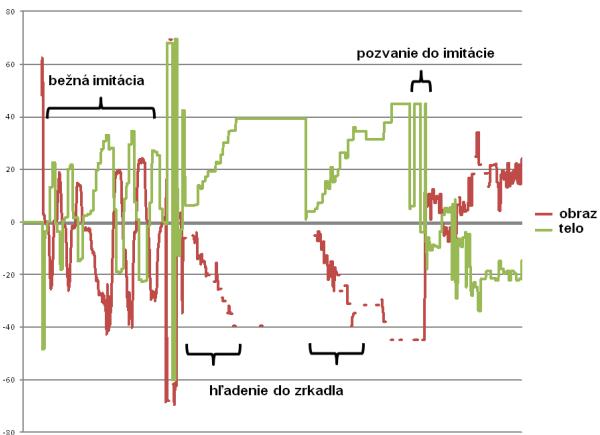
dát je veľavravný a poľahky z nich určíme, kedy sa robot díval do zrkadla (Obr. 14).



Obr. 12: Imitácia



Obr. 13: Robot sa vidí v zrkadle



Obr. 14: Záznam modelov videného a vlastného tela (os x predstavuje čas, os y uhol v stupňoch) (úplnejšia čiara je model vlastného tela, miestami prerušovaná čiara je model videného tela – vlastné telo máme vždy, ale nie vždy vidíme nejaké telo a aj keď nejaké vidíme, nie vždy to rozpoznáme)

5 Záver

V tomto článku sme predviedli demonštračný systém v ktorom prebieha uvidenie sa v zrkadle, pričom povstáva z jednoduchších mechanizmov. Uvidenie sa prebehlo na základe dokonalej korelácie medzi vlastným telom a telom robota, nevyskytujúcej sa pri uvidení tela niekoho iného. Dokonca badať istú podobnosť medzi vyprodukovaným pohybmi hlavy robota pred zrkadlom a pohybmi šimpanza pred zrkadlom.

Nie všetko máme v tejto chvíli dotiahnuté do dokonalosti, nemáme napríklad vykonaný experiment, kde proti sebe postavíme dva simulátory iCubSim, ani nám nie je celkom jasné čo by v takom experimente malo vyjsť. Lebo je to situácia obdobná tomu, keď šikovný herec v prími imituje zrkadlo. Také experimenty sa robili a podarilo sa viacero ľudí oklamať, že sa dívali na svoj obraz. Tu by práve dosiaľ v podstate nepotrebný modul "invitation" mohol zohrať rozhodujúcu rolu.

Každopádne môžeme vysloviť ako hypotézu, že schopnosti uvidieť sa v zrkadle predchádzajú v prírode schopnosť uvidieť niekoho iného (a spoznať ho ako indívídum). A taktiež, že dobrým predpokladom uvidenia sa je schopnosť konštruovať z videného model analogický vlastnému telu, čo je schopnosť prekvapujúca, ale zrejme v prírode prítomná (viď zrkadliače neuróny). Iste by bolo veľmi zaujímavé skúmať, kde táto schopnosť berie a ak z niečoho povstáva, prečo nie je úplne bežná. Ako aj to, nakol'ko je vrozená – možno sme takým druhom, ktorý mal práve to šťastie, že sa jeho zraková kôra vyvinula tak, že vyprodukovala útvar podobný motorickej kôre... Tažko povedať...

Uvidenie sa v zrkadle je kognitívna funkcia, ktorú – úmerne jej zriedkovosti v prírode – vnímame ako niečo úžasné. V prvej chvíli sme preto nadšení, že ju vieme vyvolať u robota. Avšak v druhej chvíli si uvedomíme že sme ju zložili z malého počtu modulov a to modulov dosť jednoduchých. A ten úžas sa stratí.

Na záver pripomeňme, že uvidenie sa v zrkadle nie je ešte pochopením vlastnej existencie. My sme robili robota, ktorý vie, že: "Toto som ja", nie že "Ja som". Aj schopnosť prejsť dot-mirror testom sa u detí prejavuje vo veku 18 mesiacov, zatiaľ čo ako-tak rozvinutú osobnosť má dieťa zhruba až vo veku dvoch rokov.

Prezentačné video je dostupné na adrese:
www.agentspace.org/mirror/iCubSimAtTheMirror.mp4

Zdrojové kódy sú dostupné na adrese:
www.agentspace.org/mirror/iCubSimAtTheMirror.zip

Literatúra

Bay, H. – Ess, A. – Tuytelaars, T. – Van Gool, L. (2008) Speeded-Up Robust Features (SURF).

Computer Vision and Image Understanding archive.
Volume 110 Issue 3, June, 2008, pp. 346-359

Bradski, G. (1998) Computer vision face tracking for use in a perceptual user interface.

Bradski, G. (2000) The OpenCV Library. *Dr Dobb's Journal of Software Tools*.

Brooks, R. (1999). *Cambrian Intelligence*. The MIT Press, Cambridge, MA

Burian Jan (2012) Konceptuální model prostředí pro seberozpoznávajícího robota (The conceptual model of environment for self-recognition robot).
<http://docplayer.cz/4089023-Konceptualni-model-prostredi-pro-seberozpoznavajiciho-robota.html>

Gallup, G. G., Jr. (1970). Chimpanzees: Self Recognition. *Science*, Vol. 167, 86-87.

Hart, J. W. Scassellati, B. (2012). Mirror Perspective-Taking with a Humanoid Robot. In: *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI-12)*. Toronto, Canada.

Kelemen, J. (2001). From statistics to emergence – exercises in systems modularity In: *Multi-Agent Systems and Applications*, (Luck, M., Maøík, V., Štepánková, O. Trappl, R.), Springer, Berlin, pp. 281-300

Lucny, A. (2004). Building complex systems with Agent-Space architecture. *Computing and Informatics*, Vol. 23, pp. 1001-1036

Minsky, M. (1986). *The Society of Mind*. Simon&Schuster, New York

Sandini, G. – Metta, G. – Vernon, D. (2007). The iCub cognitive humanoid robot: an open-system research platform for enactive cognition. In: *50 years of artificial intelligence*, pp. 358-369, Springer-Verlag, Berlin

Takeno, J. (2008). A Robot Succeeds in 100% Mirror Image Cognition. *International Journal on Smart Sensing and Intelligent Systems*, Vol. 1, No. 4, December 2008.

13. Viola, P. – Jones, M. (2001) Rapid object detection using a boosted cascade of simple features. *Conference on Computer Vision and Pattern Recognition*.

14. The Corpora. (2011). QBO and the mirror. What if..., <http://thecorpora.com/blog/?p=844>

Replicator and ART2 Neural Networks in Text Outliers Identifications

Asmaa Salem, Gabriela Andrejková, Pavol Kozák

Institute of Computer Science, Faculty of Science

P. J. Šafárik University in Košice

Email: asmamostafa.salem@gmail.com, gabriela.andrejkova@upjs.sk, pauluskozak@gmail.com

Abstract

The intuitive definition of an outlier should be: an observation which deviates so much from other observations as to arose suspicions that it was generated by a different mechanism (Hawkins (1980)). Two models of neural networks (1) Replicator Neural Networks and (2) neural networks based on Adaptive Resonance Theory were used in our work to identify outlier sentences in texts. The texts are written in some language, using a grammar, in some genre and they present a sequence of words, sentences, sections and we find some outlier sentences in these texts. For example, outliers should show that some parts of the text were written (modified) by another author or somebody manipulated with the text. The results of both neural networks declare that they give very good answers for manipulated texts and quite good answers for original texts. In the manipulated texts, the inserted sentences were considered outliers. In the text analysis we used English and Arabic texts from recommended benchmarks.

1 Introduction

Using neural networks models for text analysis finds quite an important role and gives interesting results in their evaluation. Self-organizing maps (SOM) were used for authorship attribution and for a stylistic analysis by Neme et al. (2015) and for dissimilarities detections by Almarimi and Andrejková (2015). In Christ et al. (2005), the neural networks based on the adaptive resonance theory (ART2) were used to create summaries of texts. Replicator neural networks were used by Dau et al. (2014) for anomaly detections. The anomaly detections using competitive neural networks were studied by Barreto and Aguayo (2009).

We suppose the following idea in our research: The texts are built from sentences. Authors of the text have their own stylistics of the text and their sentences and used vocabulary are unique for them. We suppose that it is possible to find outlier sentences in the texts. To find outlier sentences we use a combination of ART2 neural networks and replicator neural networks. The common structures of the used neural networks are known, but we modified them for our problem. The texts and their sentences have to be encoded if we plan

to use them in neural networks. The processes of encoding were the same for both networks.

The paper has the following structure: In Section 2 a very short description of Replicator Neural Networks (ReNN) from sequences learning point of view is given. In Section 3 we give information on ART2 neural networks based on Adaptive Resonance Theory (ART2). Section 4 contains information on some analyzed texts and on preprocessing of texts, Section 5 contains a description of our encoding system. The following section contains experimental results and an evaluation of the results for both neural networks. In the conclusion, we summarize our results and give some plan for the following work in the area.

2 Replicator Neural Networks

Replicator neural networks (ReNN) are based on the feed-forward neural network (FFNN) models, and they are known in the literature as autoencoders (Dau et al. (2014)). They have a special numbers of neurons in layers. The number of neurons in the input layer is the same as in the output layer as in Fig. 1.

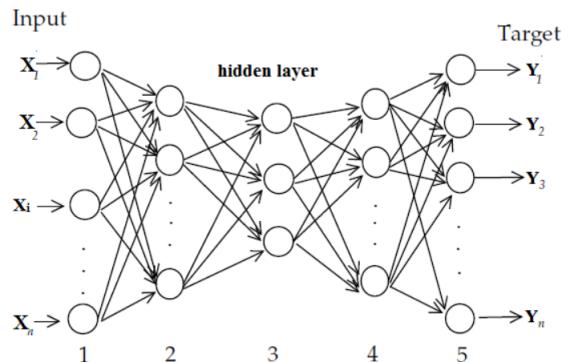


Fig. 1: A replicator neural network architecture (ReNN (2016))

Let the number of layers in ReNN be k , $k > 2$. The numbers of neurons in layers $i + 1$ and $k - i$, for $1 \leq i \leq \lfloor \frac{k}{2} \rfloor$, are equal. If these numbers are decreasing as in Fig. 1, the network is called *compression autoencoder*, if the numbers are increasing the network is *de-noising autoencoder*. All neurons work in the

classical way: some activation function is applied to the potential of neuron (the sum of its weighted inputs).

Training the ReNN. ReNN is a supervised network. The expected values on the output layer are the same as the values on input layer, it means the output of the network is compared to input values and modifications of network weights start according to the result. For the modifications of weights, methods of training FFNN can be used, for example Back-Propagation (BP) algorithm (Hertz et al. (1991)).

3 ART2 Neural Networks

The ART2 neural networks belong to the class of unsupervised and competitive learning algorithms. Adaptive resonance theory (ART) is a theory developed by Stephen Grossberg and Gail Carpenter on aspects of how the brain processes information (Carpenter and Grossberg (2014)). The structure of ART2 network is in Fig2. The network is mainly used for building of clusters.

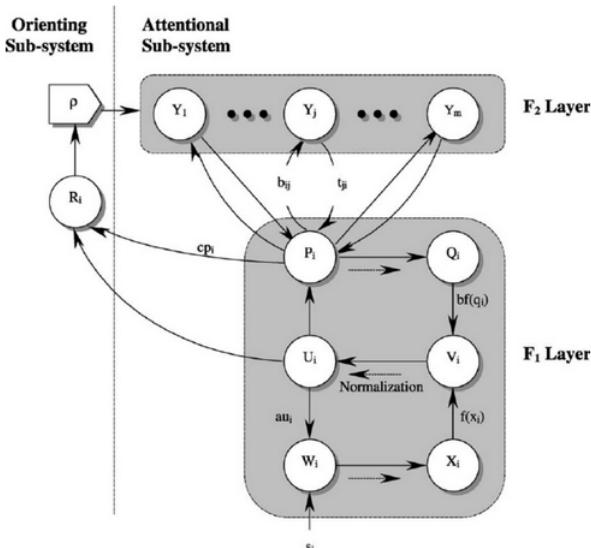


Fig. 2: ART2 Structure, ART2 (2015)

The learning algorithm:

1. initialize the network and set all parameters
2. while input exists, do the steps 2.1. – 2.4.
 - 2.1. read input
 - 2.2. update activations in the input F1 layer
 - 2.3. while not founded winner clustering node, do the following steps:
 - 2.3.1. find a node in the output F2 layer with the highest activation

2.3.2. update activations of input layer for a comparison of conformity

2.3.3. if the vigilance is OK, the neuron is the winner

2.3.4. if the vigilance is not OK, set a reset signal to the neuron

2.3.5. if all nodes in the output layer have the reset signal set, create in the layer F2 a new neuron, which will be the winner neuron

2.4. modify the weights of the winner neuron

3. end

The update processes of weights in the steps (2.2.) and (2.3.2) are quite complex and they can be found in Carpenter and Grossberg (2014). We used classical model with the length of input prepared according to the length of the average length of sentences in texts.

4 Data Preparation, Sentences of Texts

In our texts, we work with

- $\Gamma_A, |\Gamma_A| = 28$, Arabic and $\Gamma_E, |\Gamma_E| = 26$ English alphabet; Digits={1, 2, 3, 4, 5, 6, 7, 8, 9, 0}.
- V_T - a finite vocabulary of different words in the alphabet Γ presented in the alphabetic order; $|V_T|$ - the number of words in the vocabulary V_T ; each text T has its own vocabulary V_T ,
- T - a text, $|T|$ is the number of sentences in T .

We use English recommended texts from benchmark CorpusEnglish (2011) and Arabic texts from CorpusArabic (2011) in the text analysis. In Table 1 we describe some information about two Arabic and two English texts. The texts AC and EC were constructed as a combination of two different texts (important for an illustration of our outlier analysis). The position of the combinations in both texts are shown in the results of the analysis.

We will use the texts as sequences of sentences. Each data set is created from all sentences of some text. If the previous analysis was done and we know potential outlier sentences, the data set is divided into two disjunctive sets of instances (sentences) - the set of normal instances (good sentences) and the set of potential outlier instances (bad sentences). If we do not know the set of potential outliers, then it is possible to prepare the set of potential outliers using random choosing and the rest of sentences will be in the set of normal instances. Then we prepare:

- a training set - by random sampling of the instances from the normal set;

- a validation set - by random sampling of the instances from the normal set (excluding the ones used in training);
- a test set - by random sampling of the instances from the normal set (excluding the ones used for training and validation) and a number of instances (or all) from the potential outliers set;

Tab. 1: Statistics of two English and two Arabic texts, the number of words by length for 1 – 10. L_mx_sen means the maximal length of the sentences.

	Name of Texts			
#	A1	AC	EH	EC
# words	40 236	20 011	42 201	3 734
# letters	162 843	83006	209 072	15 625
≠ words	14 961	8423	6241	1142
by length				
1	86	62	828	254
2	5399	2701	7024	729
3	8836*	4357*	9498*	914*
%	21.96	21.77	22.50	23.54
4	9793	4548	5325	657
5	6758	3348	4471	380
6	4314	2217	3561	341
7	2341	1384	3743	233
8	919	548	2506	152
9	241	171	1846	116
10	63	50	1578	70
11	5	7	803	31
12	1	3	508	20
L_av_sen	13	14	26	18
# sentenc	3289	1559	1268	190

The preprocessing of texts is done in the following steps:

1. remove titles of sections; the titles of sections, references and captions to figures are not normal sentences;
2. transform capital letters to small letters;
3. remove undesirable symbols' - () ' [] : ; , “;
4. remove undesirable words - the, of, and, to, in, a, by, that, this, an, these; these words are not important for our analysis.
- Remark: Not removed words - a.m., p.m., mrs., ms., mr., dr., no.
5. convert uppercase to lowercase; the meaning of the words will not change.
6. encode of the words;

5 Encoding of texts

The encoding of a text can be done by different methods:

1. using frequencies of words divided by the total number of the words in the text; in this case different words have different code, but words with the same frequencies have the same code;
2. using the lengths of words divided by the length of the longest word; in this case texts lose a compatibility of words and all words with same length have the same code;
3. using syntactic analysis of sentences from grammars point of view; the method should be very successful but it is very time consuming.

In our experiments, we used the first method of encoding. If the word s has its frequency F_s in the text T and \max_{F_s} is the maximal frequency of words in T then the code of the word s is

$$c_s = \frac{F_s}{\max_{F_s}}, \text{ or } c_s = \frac{F_s}{|V_t|} \quad (1)$$

The first possibility gives codes with higher numbers but still from the interval $(0, 1)$.

6 Evaluation of the developed method

6.1 System of the processing

The process of an outlier detection was done in the following steps:

1. ART2 neural network was used to classify all sentences in a text into classes. Similarly encoded sentences belong to the same class. It is supposed that they were written by the same author. The classes with the number of sentences higher than some threshold br (the number is an experimental parameter in our experiments) are supposed not outlier sentences. These all sentences belong to the set of good sentences S_g . The rest of the sentences are potential outliers and belong to the set of bad sentences S_{po} .
2. ReNN is used for analysis of sentences that belong to S_{po} . The training set S_{tr} is prepared from the set S_g . The validation set S_v is prepared from both sets S_{po} and S_g , $S_v \cap S_{tr} = \emptyset$. As a testing set we used the set S_{po} and some sentences from S_g .

6.2 The evaluation method

The main criteria for the evaluation of ReNN work is the error of computation given by (2)

$$err_l = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - o_{ij}^l)^2, \quad (2)$$

where m is the number of sentences in the training set (or validation, or testing set), n is the length of sentences (γ * the average length of sentences in texts, $1 \geq \gamma \leq 2$), x_{ij} is the encoded input sentence, o_{ijl} is output of ReNN in l -th iteration.

The threshold for the evaluation of outlier sentences is given by percentage of the validation error that is computed for the validation set after the training ReNN.

6.3 Experimental results

We illustrate the results of the developed system processing for two different texts in both languages, two of analyzed texts (one from each language) are combined from two different texts, and a position of combinations or positions of inserted sentences are known too.

Let $|T|$ be the number of sentences in the text T . The set of all sentences is input to the ART2 network. ART2 creates two subsets of sentences: (1) a set A of **good** sentences in strong categories (the category is strong if it has more than $0.07*|T|$ sentences, and (2) a set B with all remaining **bad** sentences.

The training set for ReNN network is created from sentences in A , it is used 60% sentences. 30% belongs to a validation set and the rest 10% belongs to the testing set together with all sentences from B .

6.3.1 Parameter values of for ART2 and results

- n - the length of input; the maximal lengths of sentences in texts were higher than 50; the average lengths are written in Table 1; γ * the average length of sentences in texts (the value about 30); in the experiments $\gamma = 1.1$;
- a, b - weights from neurons U, Q to neurons W ; default value is 10;
- c - standard value 0.1; d - standard value 0.9;
- e - small constant for expressions with divisors, to solve division by 0;
- ρ - a vigilance parameter; it determines a minimal coincidence of input to a category for an assignment; according to the value ρ are build categories;
- α - constant for modification of the speed of learning, the standard value is 1;
- θ - default value is $\frac{1}{\sqrt{n}}$;

6.3.2 Parameter values for ReNN and results

- n - the length of input, the number of neurons in input and in output layer; γ * the average lengths of sentences in texts, it is the same value as for ART2;

- the number of neurons in the first and third hidden layers; 0.6* the length of input; the number of neurons in the second hidden layer: 3 neurons ;
- activation functions of neurons according to layers, $f_i(x)$ is the activation function for i -th layer:

$$f_1(x) = f_2(x) = f_4(x) = \tanh(x) \quad (3)$$

$$f_3(x) = \frac{1}{2} + \frac{1}{4} \sum_{j=1}^3 \tanh[(100(x - \frac{j}{4})] \quad (4)$$

$$f_5(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

6.3.3 Analyzed texts - information

Information on four types of used texts is in Table 1. The Arabic text A1 and English text EH are parts of original texts from benchmarks, Arabic text AC is combined from two different texts (half and half) and English text EC contains some inserted sentences from the other text in the positions: 1, 2, 3, 17 34 77 78 79 80 180 181 182 183.

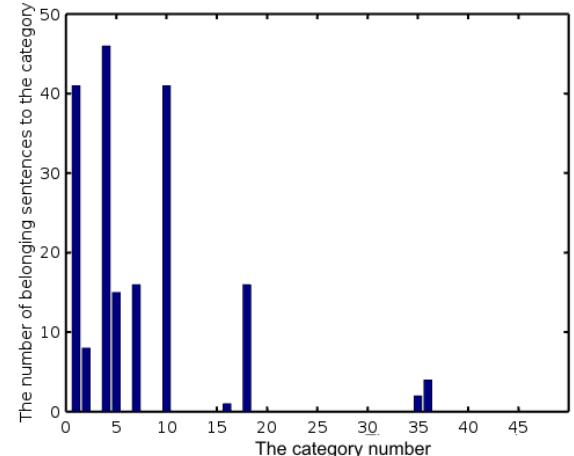


Fig. 3: The number of categories in the text EC with their capacity. The biggest category 4 has 46 sentences. The categories with one or two sentences are not visible.

Results for the text EC described in Table 3 and characterized by graphs in Fig. 3 and Fig. 4. The graphs illustrate the clusters (categories) computed by ART2 neural network. In the clusters, there are concentrated the sentences with the similar codes.

According to results partially given in the Table 3, we have the analysis of the combined text EC. The size of the testing set was 65 sentences. The set was prepared in random sequence of 10% of sentences from the set A and the rest sentences from the set B . In Table

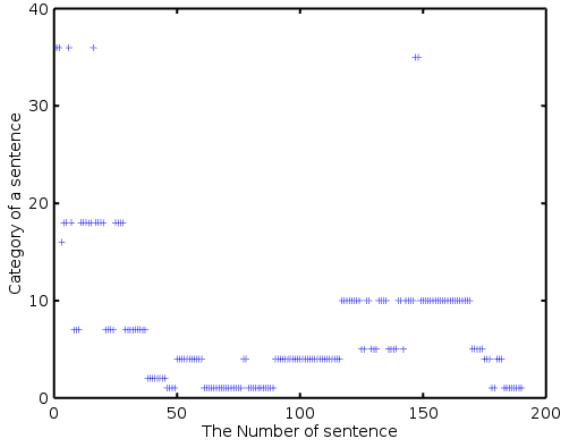


Fig. 4: The analysis of the text EC according to ART2 network. In the graph, there are plotted the numbers of sentences in categories.

Tab. 2: Evaluation of the combined text EC for the testing set with 65 testing sentences. We show only the results for inserted sentences and some other types of results for the characterization. The value of the used validation error is 0.0505. In the testing set, the thresholds for classification were 20%, 40% and 60%.

ART2	ReNN			% of results EC text
	20%	40%	60%	
0	0	0	0	27.69
0	0	0	1	13.85
0	0	1	1	10.77
0	1	1	1	26.15
1	0	0	0	6.15
1	0	0	1	9.23
1	0	1	1	0
1	1	1	1	6.15

2, we can see a percentage evaluation of all sentences using both networks.

The recognition of a sentence as a **good** sentence is written by 1 and the recognition of a sentence as a **bad** sentence is written by 0. In the recognition of ReNN were used 3 thresholds for the recognition (20%, 40% and 60% of the validation error).

The bad sentences are potential outliers, it means that their analysis is more important. 52.31% of sentences were classified as **bad** by ART2 and as bad sentences by ReNN too. 26.15% sentences, ReNN recognized as **good** sentences. The rest 22.23% sentences ART2 classified as good but 6.15% of them ReNN classified as **bad** sentences.

The results for the original EH text are different. The test set had 144 sentences, 13.19% of them were

Tab. 3: Evaluation of the combined text EC for the testing set with 65 testing sentences. In the Table we show only the results for inserted sentences and some other types of results for the characterization. The bold numbers are the numbers of inserted sentences - real outliers in the text.

# sent.	ART2 result	ReNN error	Evaluation to val. error		
			20%	40%	60%
1	0	0.1075	0	0	0
2	0	0.0589	0	0	1
3	0	0.1047	0	0	0
4	0	0.0547	0	0	1
5	1	0.0192	1	1	1
6	0	0.1177	0	0	0
7	0	0.0494	0	0	0
17	0	0.0239	0	1	1
29	0	0.0090	1	1	1
30	1	0.1078	0	0	0
40	0	0.0659	0	0	1
41	0	0.1556	0	0	0
44	0	0.0993	0	0	0
46	0	0.1014	0	0	0
50	0	0.1070	0	0	0
61	1	0.0081	1	1	1
77	0	0.0251	0	1	1
79	0	0.0247	0	1	1
97	1	0.0115	1	1	1
100	0	0.1066	0	0	0
115	0	0.0686	0	0	0
147	0	0.1035	0	0	0
166	0	0.0724	0	0	0
169	0	0.0092	1	1	1
170	0	0.1504	0	0	0
171	0	0.0177	1	1	1
175	0	0.0704	0	0	0
178	0	0.1137	0	0	0
179	0	0.0152	1	1	1
180	0	0.0073	1	1	1
181	0	0.0674	0	0	1
183	0	0.0080	1	1	1
186	1	0.0933	0	0	0
187	1	0.0075	1	1	1

bad sentences and only 1 sentence was classified as **bad** by ART2 and **good** by ReNN. 8 sentences were classified as **good** by ART2 and as **bad** by ReNN. Results for Arabic texts are quite different. The testing set for the text AC had 224 sentences, ART2 recognized 19.64% from them as **bad** sentences and ReNN gives a different

result in 40.90%. The difference in results could be due to a different structure of Arabic sentences and probably it is necessary to use different parameters.

7 Conclusion

In the paper we developed the system with two types of neural networks for detections of outlier sentences in some text. The system was tested on a small number of English and Arabic texts and it is capable to find potential outlier sentences and call attention to them. But it needs to analyze more texts.

As the future work we plan to do many statistic results to evaluations of bigger text sets and to analyze the system with modified parameters in both languages. The next plan is to use different encodings of texts for both networks and put more information to codes.

Acknowledgment

The research is supported by the Slovak Scientific Grant Agency VEGA, Grant No. 1/0142/15.

References

- Almarimi, A. and Andrejková, G. (2015). Discrepancies detection in arabic and english documents. *AC-SIJ Advances in Computer Science: an International Journal*, 4(2322-5157):69–75.
- ART2 (2015). https://www.researchgate.net/figure/222517262_fig3_fig-3-typical-architecture-of-art2-neural-network.
- Barreto, G. and Aguayo, L. (2009). Time series clustering for anomaly detection: Using competitive neural networks. *Proceedings WSOM 2009, LNCS(5629)*, pages 28–36.
- Bensalem, I., Rosso, P., and Chikhi, S. (2013). A new corpus for the evaluation of arabic intrinsic plagiarism detection. *CLEF 2013, LNCS 8138*, pages 53–58.
- Carpenter, G. A. and Grossberg, S. (2014). Adaptive resonance theory. *Encyclopedia of Machine Learning and Data Mining*, 1:22–35.
- Chomsky, N. (2005). Three factors in language design. *Lingistic Inquiry*, 36(1):1–22.
- Christ, R. E. R., Basani, T., Nievola, J. C., and Silla, C. N. (2005). The use of art2 to create summaries from texts of different areas. <https://www.researchgate.net/publication/242514338>, pages 1–6.
- CorpusArabic (2011). King saud university corpus of classical arabic. <http://ksucorpus.ksu.edu.sa>
- CorpusEnglish (2011). pan-plagiarism-corpus-2011, <http://www.uniweimar.de/en/media/chairs/webis/corpora/pan-pc-11/>.
- Dau, H. A., Cieselski, V., and Song, A. (2014). Anomaly detection using replicator neural networks trained on examples of one class. In *Proceedings of the SEAL, LNCS 8886*, pages 311–322.
- Hawkins, D. M. (1980). *Identification of Outliers*. Chapman & Hall, 5 edition.
- Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley.
- Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing*. Prentice-Hall, 1 edition.
- Neme, A., Pulido, J. R. G., Muñoz, A., Hernández, S., and Dey, T. (2015). Stylistics analysis and authorship attribution algorithms based on self-organizing maps. *Neurocomputing*, pages 147–159.
- ReNN (2016). Replicator nn, <https://blog.acolyer.org/2016/06/23/ai2-training-a-big-data-machine-to-defend/>.
- Stamatatos, E. (2010). A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, pages 538–556.

Auditory spatial discrimination with visual vs. auditory attentional cueing

R. Šebeňa¹, B. Hrebeňárová², N. Kopčo³

¹Department of Psychology, Faculty of Arts, PJ Safarik University, Moyzesova 9, 040 59 Košice,

² Faculty of Medicine, PJ Safarik University, Trieda SNP 1, 040 11 Košice,

³Institute of Computer Science, Faculty of Science, Jesenná 5, 040 01 Košice

rene.sebena@gmail.com, nkopco@gmail.com

Abstract

A previous study found an enhancement of auditory spatial discrimination ability when the listener's gaze was directed towards the auditory stimulus (Maddox et al., 2014). Here, we examined whether directing spatial auditory attention also affects this cross-modal enhancement when using realistic spatial simulation. Listeners made a judgment about the relative positions of two click-trains following a visual or auditory cue, while fixating on a neutral location. Results show that 1) subjects performed better when visual cue was used, and 2) auditory cue presented from incongruent location resulted in deteriorating performance. These results suggest a complex interaction between attentional and eye-gaze control mechanisms in auditory spatial representation.

1 Introduction

Objects and events in the real world are made up of multimodal sensory attributes.

Our nervous systems process information from different sensory modalities independently, and this information from our senses is at some point combined into one perceptual experience. Perception is a multisensory process where sensory information is integrated both within and across different sensory modalities. Some studies have shown that auditory and visual stimuli can be integrated by bimodal cells, exhibiting spatially overlapping auditory and visual receptive fields. Such neurons have been found in the early sensory cortical areas such as superior colliculus (e.g. Lakatos et al., 2007, 2008; Kayser et al., 2009) and recent study found that multi-sensory effects have been shown to occur in primary sensory areas as well (Lemus et al., 2010).

Multimodal activation has also been found in the human parietal cortex (Bremmer et al. 2001; Bushara et al. 1999, 2003; Cusack et al. 2000; Warren et al. 2002) and Intraparietal sulcus in the areas commonly referred to as LIP (lateral intraparietal sulcus bank) and MIP (medial intraparietal sulcus bank). Neuron cells in this area have been found to be sensitive to the locations of both visual and auditory stimuli (O'Daniel et al.,

2005; Ben Hamed et al. 2001, 2002; Cohen et al. 2004; Gifford and Cohen 2004; Cohen and Andersen 2000).

Information from one sense has the potential to influence how we perceive information from another. For example irrelevant visual stimulus can affect the detection of an auditory stimulus (Lovelace et al., 2003) as well as the perceived loudness (Odgaard et al., 2004).

Attention facilitates selection of objects, events, or spatial regions in complex scenes. Very few studies focused on the effect of attention on sound localization. Even fewer studies looked at whether the effect is modality-dependent. Only a few previous studies asked whether directing automatic (exogenous, involuntary, stimulus-driven) or strategic (endogenous, voluntary, goal-driven) attention by an auditory cue can improve sound localization (Spence & Driver, 1994; Sach et al., 2000; Kopco et al., 2001). The result showed that cueing caused improvements in reaction times (Spence & Driver, 1994), but small (Sach et al., 2000) or no (Kopco et al., 2001) improvements in localization accuracy. Possible reasons were that tested SOAs were too short to orient attention and that auditory cue is not efficient because audition is not primarily a spatial modality.

A recent behavioral study demonstrated enhancement of auditory spatial cue discrimination ability when the listener's gaze was directed towards the auditory stimulus (Maddox et al., 2014). However, such an effect has only been demonstrated for simplistic binaural cues (interaural time and level differences). In the current behavioural study, we are expanding the findings of this paper by utilizing head related transfer functions (HRTFs) and by examining whether spatial auditory attention also affects this cross-modal enhancement.

We hypothesized that automatic attention will improve performance for valid trials and little effect or decrease in accuracy for invalid trials. We also hypothesized that effect of automatic attention will be modality dependent. Specifically based on Maddox paper, we assumed that there will be better performance in valid visual cue trials compared to valid auditory cue trials. The aim of this study is also to obtain behavioral data for electrophysiological analysis

of auditory event related potential changes in cortical brain areas.

2 Methods

13 subjects (9 male, aged 20 - 38 years) participated in the two-session experiment. All participants were without any known hearing deficiencies. Some initial practice trials on each of the different experimental conditions were given prior to data collection. All provided written informed consent as approved by the PJ Safarik University in Kosice.

Auditory and visual stimuli were generated using Matlab (Mathworks, Natick, MA). The experiment was controlled using Matlab with the Psychtoolbox 3 extension (Brainard, 1997). Sound stimuli were presented using Etymotic Research (Elk Grove Village, IL) ER-1 insert headphones connected to a Datapixx system (VPixx Technologies, Saint-Bruno, QC). During the experiment subjects sat in a sound-treated booth (Eckel Laboratories). The experiment consisted of visual cue and auditory cue trials. On each trial there was target consisting of two sounds presented from slightly different locations and listeners task was to discriminate the direction of the target location change.

Visual cue and auditory cue trials

Visual cue trials consisted of a 100 ms white dot presented in horizontal plane at either 0 degrees or +/- 25 degrees on the computer screen. Subjects were instructed to fix their gaze at +12.5° or -12.5° (balanced across trials) position during the whole trial, to pay attention to cue and to expect target stimulus from the same position as the presented cue. At 800 ms an auditory target was presented either about 0 or +/- 25 degrees laterality ipsilateral with the fixation point through insert earphones. The auditory target consisted of two, 100 ms click trains played successively. The first click train was presented at 0 or +/- 25 degs, and the second train at a location slightly shifted (4.2° for central and 8.4° for lateral position); relative to the location of the first one. The subject's task was to respond, after the auditory stimuli disappeared, if the target moved left or right (using 1 or 2, respectively, on the keyboard).

Auditory cue trials were similar to the visual primer trials, consisting of an auditory target being played at 800 ms lasting for 200 ms, however the 100 ms click train at 170Hz served as an auditory cue at the beginning of each trial. Subjects were instructed to pay attention to cue, to expect target stimulus from the same position as presented cue and respond in the same manner. Schematic of the trial sequences is presented in Figure 1.

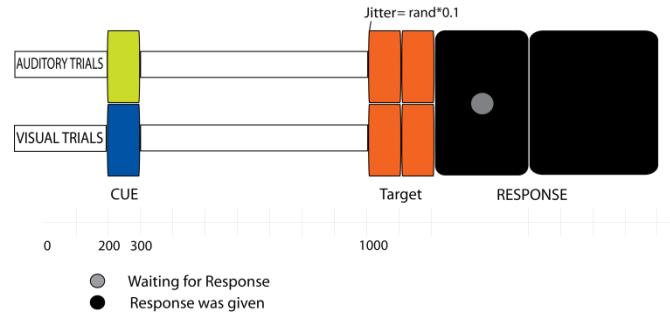


Fig.1: Schematic of the trial sequences and the two experimental conditions

Conditions

In 'matched' trials (50%), the target stimuli were presented at the same location as the white dot. In 'mismatched' trials (50%), the target was played in the opposite location of the visual cue (0-deg cue for 25-deg target or 25-deg cue followed by a 0-deg target). In the auditory cue conditions, 'matched' trials consisted of the auditory cue and the probe occurring at the same location. 'Mismatched' auditory cue trials had the auditory cue and target occurring in opposite locations (Figure 2).

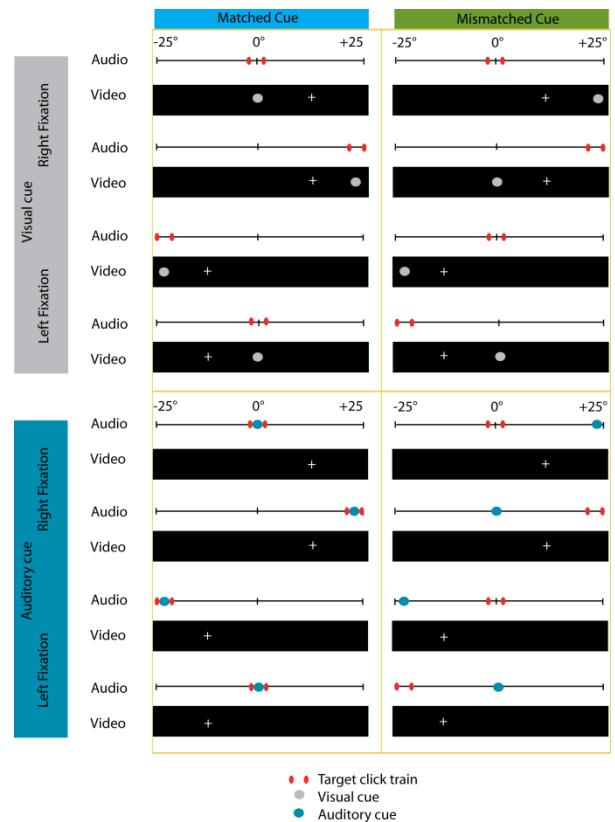


Fig. 2: Experimental conditions

Data analysis

For most of the conditions, there was left-right symmetry in results. Therefore, data collected with fixation point on the left were mirror-flipped and

combined with the data collected with fixation point on the right. And, unless stated differently, the data are presented as if the fixation point was at 12.5°.

3 Results

Data were analyzed for auditory and visual trials and for ‘Match’ versus ‘Mismatch’ conditions and for central and lateral target position. The percentage of correct responses for all participants and their means were computed (see Figures 3 and 4).

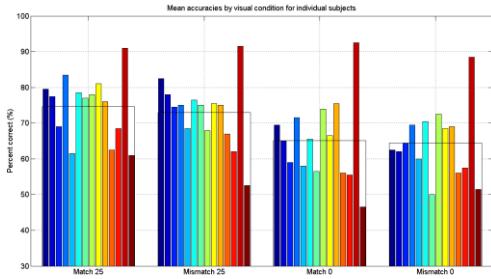


Fig. 3: Percentage of correct responses for individual subjects (color bars) and across-subject mean (open bar) for visual trials and for match, nonmatch cue (M vs. N in labels) and for position of target (0 vs. 25 in labels)

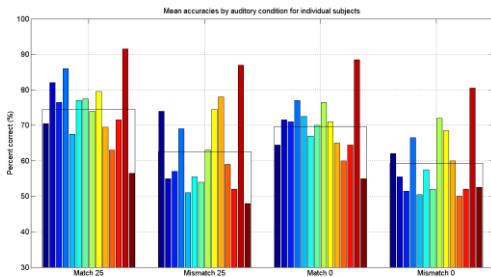


Fig. 4: Percentage of correct responses for individual subjects and their means for auditory trials and for match, nonmatch cue and for position of target

To identify significant differences between experimental conditions, repeated measures ANOVA was performed on RAU-transformed % correct data (Tab. 1).

	df	F	pValue
Cue	1,13	7.32	0.018
Position	1,13	35.97	< 0.001
Fixation	1,13	3.93	0.069
Matching	1,13	47.13	< 0.001
Cue x Matching	1,13	12.62	0.004
Cue x Position	1,13	6.62	0.023
Fix. x Position	1,13	5.87	0.031
Cue x Fix x Match	1,13	3.84	0.072

Tab. 1: rANOVA table for testing within-subjects effects

The ANOVA showed a significant effect of the Cue type ($F(1,13) = 7.3$; $p < 0.05$), indicating that in auditory cue trials, participants performed worse than in visual cue trials. There was also a main effect of Position ($F(1,13) = 35.9$; $p < 0.001$) resulting in a less pronounced decrease in performance in lateral than central position. ANOVA also indicated a significant Cue by Position interaction ($F(1,13) = 12.6$; $p < 0.01$) (Figure 5). In lateral position performance was better for visual cue compared to auditory cue.

As hypothesized we also found a main effect of Matching ($F(1, 13) = 47.13$; $p < 0.001$) resulting in a significantly better performance for matched than mismatched trials. There was also a significant Cue x Matching interaction. The difference between ‘Match’ and ‘Nonmatch’ condition for visual cue trials was not significant ($F(1,13) = 1.5$; $p = 0.24$), but for auditory cue trials this difference was significant ($F(1,13) = 28.9$; $p < 0.001$), the performance was better for ‘Match’ experimental conditions (Figure 6).

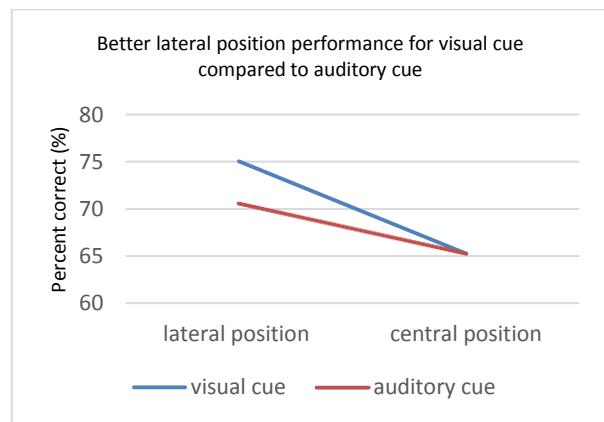


Fig. 5: Cue by target Position interaction

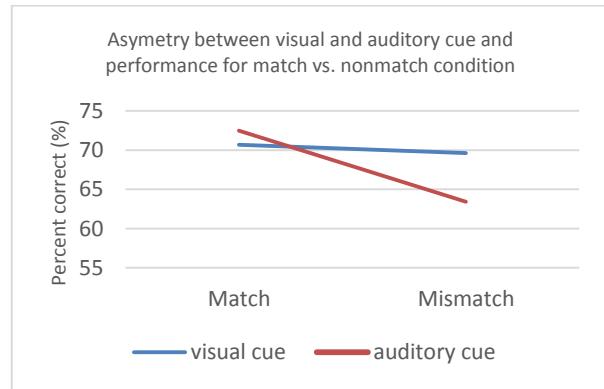


Fig. 6: The difference between ‘Match’ and ‘Nonmatch’ condition for visual and auditory cue trials

Finally we found no significant main effect of Fixation ($F(1,13) = 3.93$; $p = 0.07$). The performance was relatively the same for left fixation and right fixation. On the other hand there was a significant Fixation x Position interaction. In right fixation

performance was better for lateral presented target stimuli (Figure 7). Cue by Fixation by Matching interaction was not significant.

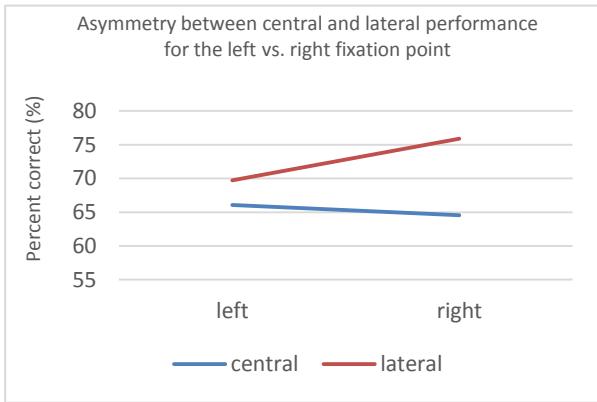


Fig. 7: Fixation by Position interaction

4. Discussion and Conclusions

Performance was better when visual cue was used
We aimed to compare subjects' performance when visual cue versus auditory cue was used. Based on our findings it can be concluded that subjects performed better when visual cue was used. This is in line with some other studies showing that visual cues help auditory perception by guiding attention to discriminate target either by enhancing sounds near the threshold of audibility when the target is energetically masked or by enhancing segregation when it is difficult to direct selective attention to the target (Varghese et al., 2012). It seems that visual cues can provide perceptual benefits helping listeners focus selective attention on the target.

Auditory cue presented from incongruent location resulted in deteriorating performance
In our experiment all trials had either 'Matched' or 'Nonmatched' cues. We found that only for auditory cue trials performance was better when cue was matching the target position. This result is surprising, partly in contrast to the previous experiments which shown intelligibility and discrimination benefits of knowing where to listen (Best et al, 2007; Maddox et al., 2014). Those gains may come from facilitated selective attention (Mesgarani, Chang, 2012).

This finding is in opposite with Maddox paper, who found for directional auditory primers no benefit for performance compared to uninformative and better performance in spatial discrimination visual directional trials than in visual uninformative trials for ILD at both the center and side positions and for ITD only when stimulus was located on the side.

An important difference between this study and the previous studies is that here only automatic spatial auditory attention was examined since the cue was only

informative at 50% of trials, thus making it unlikely that the subjects would use it to direct their strategic attention. However, it is possible that some strategical attention was engaged. Additional experiments need to be performed to distinguish between these two options.

Better performance in lateral than central position

We also found significant main effect of position with better performance in lateral than central position. This finding is in opposite with Maddox who found center performance better than side performance. In our case lateral position was much easier to discriminate due to more spatial difference between two target stimuli (8.4°) compared to 4.2° difference for central stimuli as is obvious from initial practice no cue trials on each experimental conditions which were given prior to data collection (Figure 8).

We also observed an asymmetry between central and lateral performance for the left vs. right fixation point (Fig. 7). This asymmetry is likely due to the use of non-individualized HRTFs which might have been better matched to the individual subjects' HRTFs on the right-hand side compared to the left-hand side. However, as shown in Fig. 8, which shows the no cue baseline performance measured prior to the experiment, this performance was well matched across the locations in the experiment.

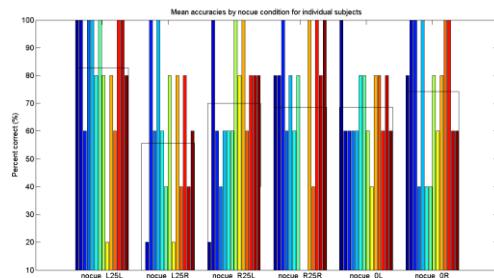


Fig. 8: No cue practice trials on 6 experimental conditions. Each group of bars corresponds to performance of individual subjects (color bars) and across-subject mean (open bar) for one combination of target positoin (L25, R25 or 0) and direction of shift.

In lateral position performance was better for visual cue compared to auditory cue

Maddox study confirmed that gazing leftward would shift the receptive field to the left, resulting in better discrimination of the left-lateralized sounds (Maddox et al., 2014). It is not clear how to relate this result to the current results, given that the subjects were specifically instructed not to move their eyes in the current experiments (and we monitored eye position using electro-oculography). It is possible that the presentation of visual cue or auditory cue induced an automatic orienting response or response planning, which then affected performance, in particular for the non-matching cues in the auditory condition.

Future studies

In the future, with regards to these studied experimental conditions and data, we plan to conduct d-prime analysis and analysis of auditory event related potential changes in cortical brain areas.

Acknowledgement

This work was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex Environments. Barbara Shinn-Cunningham, Lenny Verghese and Hannah Goldberg provided an early version of the experimental code for this study.

References

- Best, V., Ozmeral, E.J., and Shinn-Cunningham, B.G. (2007). Visually guided attention enhances target identification in a complex auditory scene. *J. Assoc. Res. Otolaryngol.* 8: 294–304
- Ben Hamed, S., Duhamel, J.R., Bremmer, F., and Graf, W. (2001). Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. *Experimental Brain Research.* 140: 127–144
- Ben Hamed, S., Duhamel, J.R., Bremmer, F., and Graf W. (2002). Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze. *Cerebral Cortex.* 12: 234–245
- Bremmer, F., Schlack, A., Duhamel, J.R., Graf, W., and Fink, G.R. (2001). Space coding in primate posterior parietal cortex. *Neuroimage.* 14: 46–51
- Bushara, K.O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., and Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience.* 6: 190–195
- Bushara, K.O., Weeks, R.A., Ishii, K., Catalan, M.J., Tian, B., Rauschecker, J.P., and Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience.* 2: 759–766
- Cohen, Y.E., Cohen, I.S., and Gifford, G.W. (2004). Modulation of LIP activity by predictive auditory and visual cues. *Cerebral Cortex* 14: 1287–1301
- Cohen, Y. and Andersen, R. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron* 27: 647–652
- Cusack, R., Carlyon, R.P., and Robertson, I.H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience.* 12: 1056–1065
- Gifford, G.W. and Cohen, Y.E. (2004). Effect of a central fixation light on auditory spatial responses in area LIP. *Journal of Neurophysiology.* 91: 2929–2933
- Kayser, C., Petkov, C.I., Logothetis, N.K., (2009). Multisensory interactions in primary auditory cortex: fMRI and electrophysiology. *Hearing Research*, doi:10.1016/j.heares.2009.02.011.
- Kopco, N., Ler, A., and Shinn-Cunningham, B. (2001). "Effect of auditory cuing on azimuthal localization accuracy," *Journal of the Acoustical Society of America.* 109, 2377
- Lakatos, P., Chen, C.M., O'Connel, M.N., Mills, A., Schroeder, C.E., (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53: 279–292
- Lemus, L., Hernández, A., Luna, R., Zainos, A., Romo, R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments? *Neuron.* 67:335-348
- Lovelace, C.T., Stein, B.E., Wallace, M.T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Research Cognitive Brain Research.* 17:447–453
- Maddox, R. K., Pospisil, D. A., Stecker, G. C., and Lee, A. K. C. (2014). Directing eye gaze enhances auditory spatial cue discrimination. *Current Biology.* 24: 748–752
- Mesgarani, N., and Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature.* 485: 233–236
- O'Daniel, A., Mullette-Gillman, Yale, E. Cohen, Y.E., and Groh, J.M. (2005). Eye-Centered, Head-Centered, and Complex Coding of Visual and Auditory Targets in the Intraparietal Sulcus, *Journal of Neurophysiology.* 94: 2331–2352
- Odgaard, E.C., Arieh, Y., Marks, L.E. (2004). Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective & Behavioral Neuroscience.* 4:127–132
- Sach, A.J., Hill, N.I., and Bailey, P.J. (2000). Auditory spatial attention using interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance.* 26(2):717-729
- Spence, C.J. and Driver. J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance.* 20(3): 555–574
- Varghese, L.A., Ozmeral, E.J., Best, V., Shinn-Cunningham, B.G. (2012). How visual cues for when to listen aid selective auditory attention. *J. Assoc. Res. Otolaryngol.* 13:359–368
- Warren, J.D., Zielinski, B.A., Green, G.G., Rauschecker JP, and Griffiths, T.D. (2002). Perception of sound-source motion by the human brain. *Neuron* 34: 139-148

Exploring the Neural Correlates of Different Interaction Types: a Hyperscanning Investigation Using the Pattern Game.

Beáta Špiláková, Daniel J. Shaw

CEITEC Masarykova univerzita
Kamenice 753/5
625 00 Brno
Česká republika
beata.spilakova@ceitec.muni.cz
daniel.shaw@ceitec.muni.cz

Abstrakt

Hyperscanning affords insight into social interaction brain dynamic by simultaneously scanning two or more individuals' brain responses while they engage in dyadic exchange. The present research aims to provide an experimental paradigm for hyperscanning research capable of delineating among two dimensions of interaction: (1) interaction structure (concurrent vs. turn-based) and (2) goal structure (competition vs cooperation). Dual-fMRI was used to scan 22 pairs of participants as they played the modified Pattern Game, where participants either compete or cooperate. Different patterns of activation between the conditions were found including insula and medial cingulate cortex in cooperation and frontal and parietal activations in competition. Turn-based condition showed supplementary motor area and frontal activations, in concurrent condition angular gyrus were activated.

1 Introduction

Humans are without doubt social creatures. We interact with ease daily, we talk to our loved ones, cooperate with our colleagues, compete with our friends on game night, we even frequently interact with complete strangers. Social neuroscience, research field primarily interested in neural mechanisms of social interactions, has revealed the neural underpinnings of many important socio-cognitive phenomena, from face processing to empathic awareness. This type of research usually involves measuring the response of one individual's brain while they evaluate social experimental stimuli. It comprises most of the current brain imaging literature and can be considered to represent "spectator science" (Hari et al., 2015) that describes a person as a detached observer, rather than actively engaged with another in some joint project (Schilbach et al., 2013). Such approach is perfectly sufficient if we are exploring the mechanisms

underlying intra-personal phenomena, such as an individual's representation of others' actions and mental states (Konvalinka & Roepstorff, 2012). It offers little or no insight into the neural mechanisms involved in mutual interaction or real-world social behaviour, however. Human social interaction is complex, dynamic and dependent on many situational factors. There is a need in social neuroscience to adopt a more interactive notion. As opposed to "spectator science" (research, where modulations of brain states triggered by the interacting partner's behaviour are neglected (Hari et al., 2015)), a new approach in social neuroscience emerged. As a mean of achieving this goal, hyperscanning emerged this two-brain approach attempts to elucidate functional relationships between two brains during mutually active interactions instead of an isolated brain processing delivered stimuli - two-person nature. Hyperscanning involves simultaneous neuroimaging of two or more individuals during interaction. It thus enables monitoring of the neural dynamics in real world social interaction and monitor activation within, but also between brains of interacting persons. So far, it has been successfully used with different imaging techniques (fMRI, EEG, fNIRS, MEG) on different paradigms like economic games (Ota, Fujii, Suzuki, Fukatsu, & Yamadori, 2001)(e.g. Astolfi et al., 2012; Billeke et al., 2014; King-Casas et al., 2005), music performance (e.g. Babiloni et al., 2012; Lindenberger, Li, Gruber, & Müller, 2009; Sänger, Müller, & Lindenberger, 2013) or different forms of verbal interaction (e.g. Baess et al., 2012; Jiang et al., 2012; Spiegelhalder et al., 2014). Hyperscanning as a technique is becoming widely popular, however, it is a relatively new method and while it has been successfully used many times, it comes with many challenges. Since it is a very complex research technique, some methodological considerations arise. This includes for example the level of acquaintanceship involved in neuronal coupling. Further it is necessary to identify, develop and optimise analytical techniques capable of exploring

interbrain effects measured in hyperscanning settings (Burgess, 2013; Hari et al., 2015). There is also a question of individual personality traits that modulate social interactions. More importantly, since the main aim of hyperscanning is to explore interaction, it is essential to characterise the precise form of interaction investigated in different experimental paradigms (Konvalinka & Roepstorff, 2012). To our knowledge, there is no research in hyperscanning that systematically focuses on several diverse properties of interaction. Here, preliminary results from an experimental hyperscanning paradigm that would consider different types of interaction and partly fill this gap are presented. To define the core properties or dimensions of interaction, I build upon a framework published by Liu and Pelowski (Liu & Pelowski, 2014). They identify three factors that shape interpersonal interaction: (1) interaction structure (concurrent vs. turn-based interaction) (2) goal structure (competitive vs. cooperative interaction) and (3) the task structure (interdependent tasks, where both individual behaviour and outcome are affected by each other vs. independent tasks, where individuals complete the task independently, while outcome (winning or losing) is determined by the other). This three-dimensional construct will provide a framework for tackling the differences that each type of interpersonal interaction brings into hyperscanning. Here, the focus is on half of them, the interdependent tasks (2 interaction structure types x 2goal structure types), since in these tasks we can better see real time dynamics of the interaction, the intertwined behaviour within the task. We modified a simple game called the Pattern Game, introduced by Decety et al. in a competition-cooperation research (Decety, et al., 2004) to match all the interaction types; and scanned pairs of participants with two identical MRI scanners while playing the game iteratively.

2 Methods

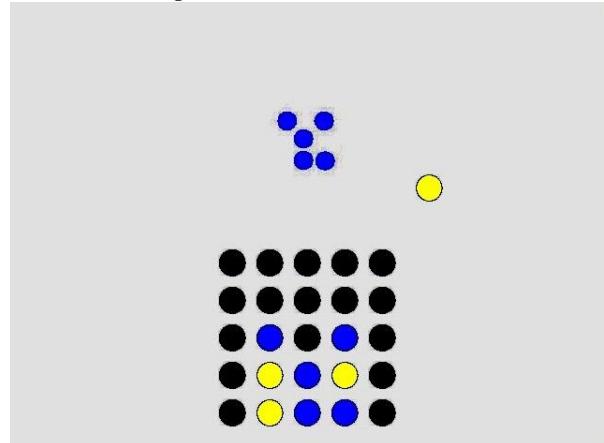
2.1 Participants

We recruited 44 participants (22 women), mainly among students of Masaryk University. They formed 22 pairs matched on gender (11 female-female and 11 male-male pairs); age (maximum age difference of 12 month, mean age difference was 6.3 months); handedness (1 left-handed female pair and 2 left-handed male pairs) and education. Mean age of participants was 22.4 years ($SD=1.9$). All participants gave their informed consent and the experiment was approved by a local ethical committee.

2.2 Modified Pattern Game

In this paradigm, each participant was assigned with a colour (blue or yellow) that remained the same

throughout the experiment. Prior to the beginning of every round, participants saw instructions that specified the role of each player: one of the participants was always the builder, the second participant was either helper or hinderer. Builder's aim was to recreate a simple 5-token pattern on a 5x5 grid (game board). The helper was instructed to help the builder recreate the pattern (cooperation condition), the hinderer's goal was to prevent builder from creating the pattern (competition condition). Participants controlled the game with simple MR compatible controllers with 4 buttons. Each participant's token automatically appeared above the game board on the designated side of the player. Participants then controlled horizontal movement of the token with two buttons (left/right) until it was positioned above the selected column. Then, after the press of the third button (down), the token fell on the last available position (as if subjected to gravity). Patterns, as well as players' roles, change every round and all of them were constructed to be impossible to recreate without the help of the other player (three supporting tokens are needed for each pattern; see Obr. 1). The control condition required one participant to recreate the pattern as well as possible (reach maximum of correctly placed tokens) and during this condition the second participant had to watch without interference. In each round, each participant had 5 tokens to place.



Obr. 1: Example of successful cooperation round in the Pattern game. Blue player is the builder and yellow is the helper.

2.3 Experimental paradigm

Prior to the experiment, all participants filled an online questionnaire; the Personality Styles and Disorder Inventory (Kuhl, Kazén, 2002). Participants of each pair were introduced to each other and instructed together. After they filled the informed consent, security questionnaire for MRI laboratory and read instructions for the Pattern Game, they both separately completed 4 practice rounds of the game. During the practice, they did not interact or play the game

together, it was constructed to simulate the actual game without the need of the co-player.

After participants confirmed their understanding of the Pattern game, they were prepared for the fMRI, where first the anatomical scans were obtained and then they played 2 blocks of 48 rounds of the Pattern Game. Each block consisted of 16 cooperation, 16 competition and 16 control rounds (always with 8 rounds where the blue player was the builder and 8 rounds where the yellow player was the builder). Each round started with the instructions displayed for 3 seconds and a 1 second fixation cross. The only difference in the first and the second block was that in the first block participants were placing their tokens alternatively (e.g. builder placed his first token than helper placed his first token, then builder placed his second token etc.; Turn-based condition); in the second block players placed their tokens concurrently (both players were placing their tokens at the same time; Concurrent condition).

2.4 Functional MRI data acquisition

For each participant, structural and functional fMRI data was simultaneously recorded with two identical 3T Siemens Prisma scanners. To ensure temporal synchronisation of the signal, an external signal generator (Siglent SDG1025) was used. This resulted in pairs of fMRI signal that had no bigger acquisition delay than 20 msec. For the purposes of localisation and co-registration, structural MR images were acquired before the functional runs (MPRAGE; TR=2300 msec; matrix=240x252x224 mm; 1x1x1 mm voxels). Blood-oxygen-level dependent (BOLD) images were acquired with a T2*-weighted echo-planar imaging (EPI) sequence with parallel acquisition (TR = 2000 msec; matrix = 68x68x34; 3x3x4 mm voxels). Slices were acquired in interleaved order. Functional imaging was performed in two runs, both comprising 570 volumes (19 minutes).

2.5 Analysis of fMRI data

For each of the subjects and each of the two time-series, data were pre-processed separately using tools from FMRIB Software Library (FSL, Jenkinson et al., 2012). Motion correction was performed using MCFLIRT with the middle volume reference. Slice-timing correction was conducted. To identify any signal related to noise sources probabilistic independent component analysis was performed using MELODIC (Beckmann, 2012), resulting 50 independent components. Artefactual components were then identified automatically with the Spatially Organized Component Klassifikator (SOCK; Bhagatgarapu, Jackson & Abbott, 2013), and signal relating to these noise components was regressed out of the time-series. Lastly, time-series was registered to native space.

This work presents preliminary results of general linear model analyses. Responses of one player to other's player successful moves were modelled (e.g. builder's response to helper's successful placing of the token). For each blue-yellow pair the condition-specific (competition, cooperation, control) correlation in BOLD signal was calculated between spatially corresponding voxels. These correlation coefficients were then converted to z-scores using Fishers-transformation. Group results presented here are familywise error corrected.

3 Results

3.1 Behavioural results

The maximum of successfully placed tokens in cooperation rounds by both builder in each block was 80 (there were 16 cooperation trials in each block, in each cooperation trial, builder placed a maximum of 5 successful tokens, $16 \times 5 = 80$). In turn-based condition, an average of successful placements in cooperation trials was $M = 74.5$ ($SD = 4.1$); in the concurrent condition the average was $M = 75.8$ ($SD = 7.1$). These numbers suggest that participants understood the task well and were able to cooperate successfully. In competition, in turn-based condition builders' average success was $M = 31.4$ ($SD = 2.8$), in concurrent condition it was $M = 36.1$ ($SD = 5.1$). As the success to the builder is significantly lower, this indicates, that participants successfully posed as hinderers in preventing builder from creating the pattern.

3.2 Builder

First; we looked at the activations in the builder's brain when the helper successfully placed a supporting token. In Turn-based condition (Fig. 1) we found bilateral activity in frontal and parietal cortex, in supplementary motor area, cerebellum and medial and anterior cingulate cortex. In Concurrent condition (Fig. 2), we saw massive activations including frontal, temporal and occipital cortex, insula, caudate nucleus, putamen, hippocampus and amygdala. Bilateral activations were also present in posterior and medial cingulate cortex. When the hinderer successfully placed a token that prevented builder from making the pattern (e.g. placed token on a position, where builder wanted to place it); during Turn-based trials (Fig. 3) builder showed activations in parietal cortex, insula and medial and anterior cingulate cortex bilaterally, left medial frontal gyrus, right superior frontal gyrus and right cerebellum. In Concurrent trials (Fig. 4) activations were present bilaterally in anterior and medial cingulate cortex.

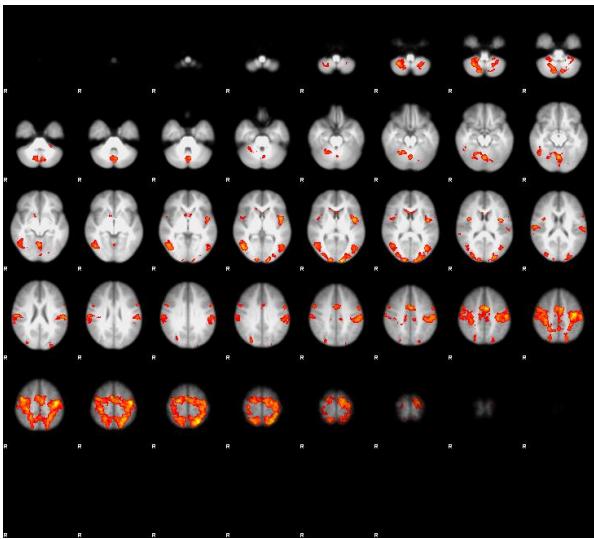


Fig. 1: Builder's reaction to helper correctly placing the token in Turn-based cooperative trials.

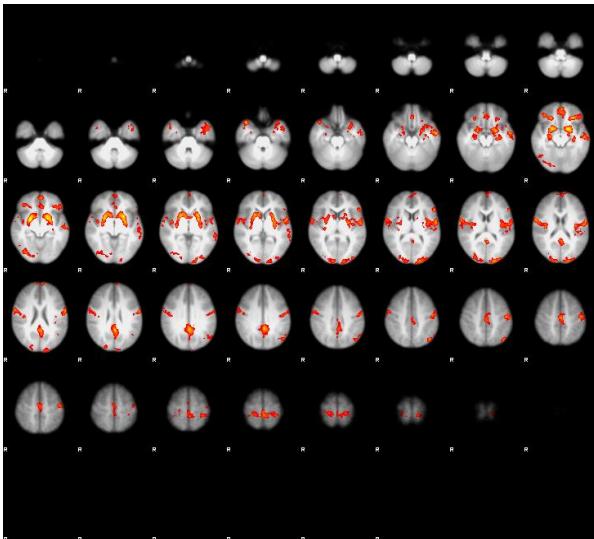


Fig. 2: Builder's reaction to helper correctly placing the token in Concurrent competitive trials.

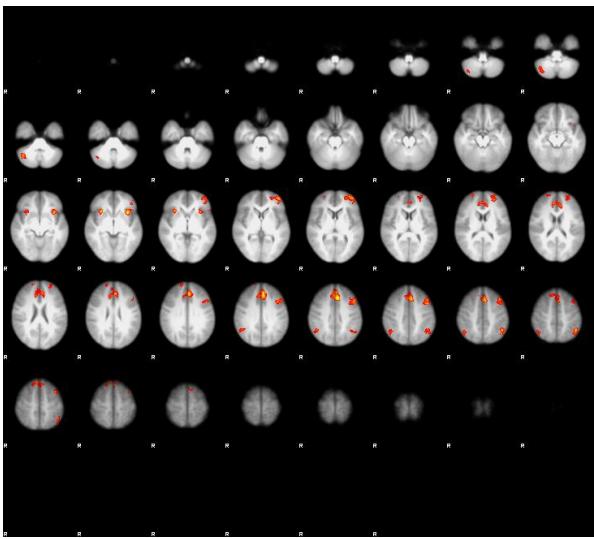


Fig. 3: Builder's reaction to hinderer placing the preventive token in Turn-based competitive trials.

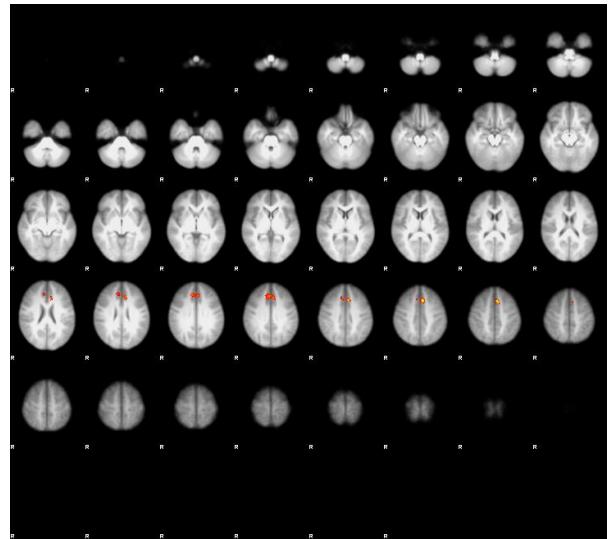


Fig. 4: Builder's reaction to hinderer placing the preventive token in Concurrent competitive trials.

3.3 Helper

When a builder placed a token on a right position in cooperation trial, in both Concurrent and Turn-based condition his co-player, the helper, showed activations in precuneus and cuneus, anterior and medial cingulate cortex, medial prefrontal cortex, right and left hippocampus and insula and also bilaterally putamen (Fig. 5 and 6).

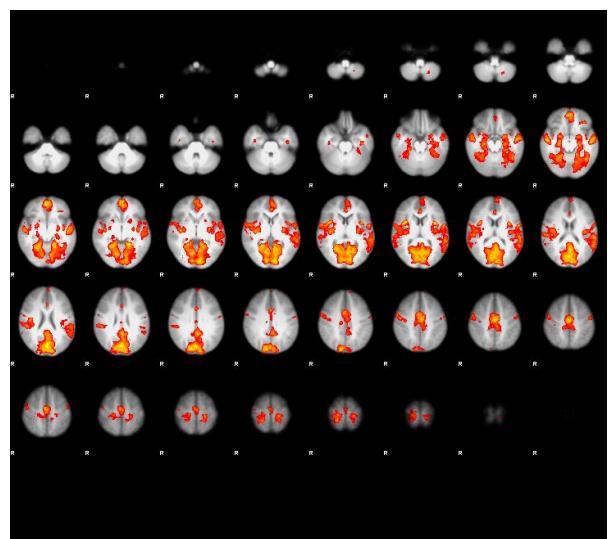


Fig. 5: Helper's reaction to builder correctly placing the token in Turn-based cooperative trials.

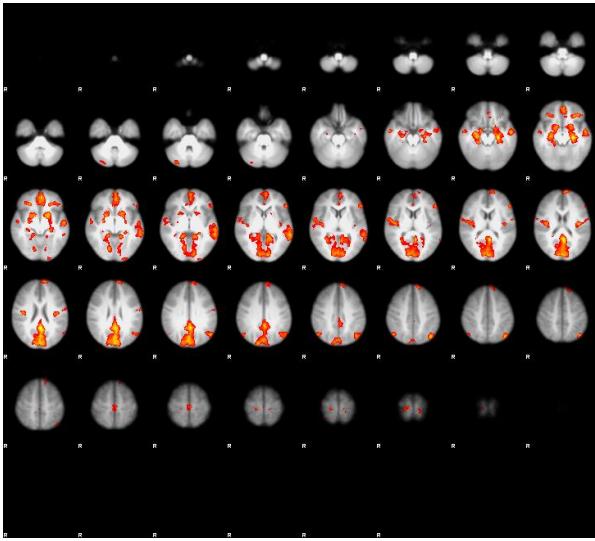


Fig. 6: Helper's reaction to builder correctly placing the token in Concurrent cooperative trials

3.4 Hinderer

Hinderer presented activations, in both Concurrent and Turn-based condition, in superior frontal lobes, parietal cortex, insula, thalamus and cerebellum when builder correctly placed a token in competition trials (Fig. 7 & 8).

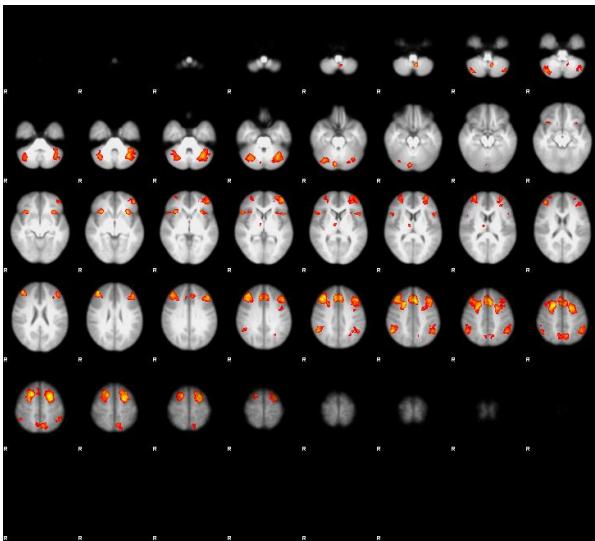


Fig. 7: Hinderer's reaction to builder correctly placing the token in Turn-based competitive trials

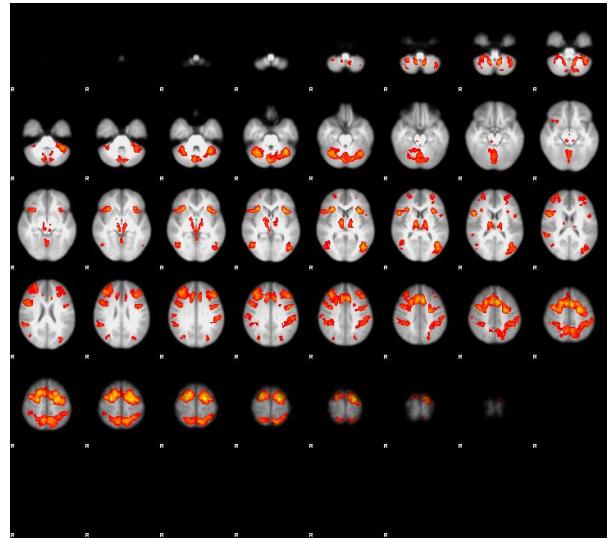


Fig. 8: Hinderer's reaction to builder correctly placing the token in Concurrent competitive trials

3.5 Turn-based and concurrent condition

In turn-based trials, where participants alternated in placing the tokens the activations were much more prominent. We found extensive activations in precentral cortex and supplementary motor areas, inferior frontal gyrus and middle and superior occipital lobes (Fig. 2). In concurrent condition activations in left medial and inferior areas and right cerebellum as well as in angular gyrus bilaterally were found (Fig. 3).

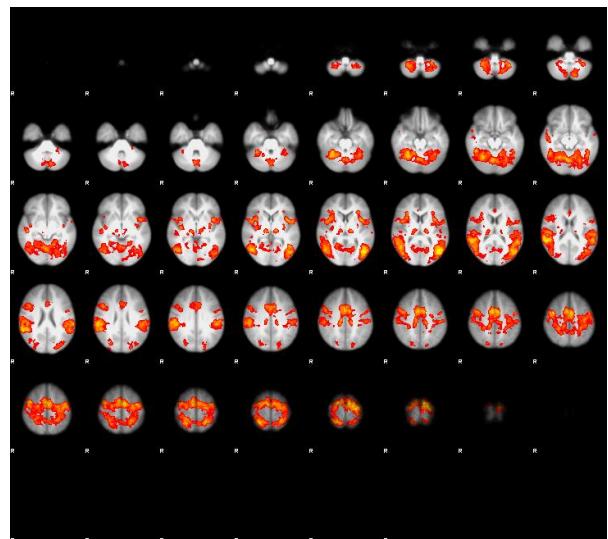


Fig. 9: Activations present in turn-based condition as opposed to concurrent condition.

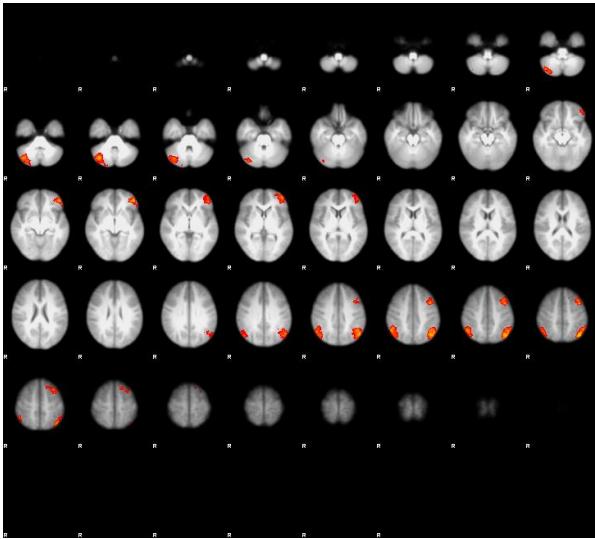


Fig. 10.: Activations present in concurrent condition as opposed to turn-based condition.

3.6 Inter-subject correlations

In all four conditions, we found correlated activity in superior frontal gyrus bilaterally. In cooperation rounds in concurrent condition we also found inter-subject correlations left medial temporal gyrus and precuneus bilaterally. In competitive rounds in turn-based condition, left precuneus showed correlated activation.

4 Discussion and conclusion

In this research, Dual-fMRI was used to scan 22 pairs of participants- each pair matched on gender, age, education and handedness- as they played the Pattern Game. In this simple interactive task, one player attempts to recreate a pattern of tokens while the second player must either help (cooperation) or prevent the first from achieving the pattern (competition). Each pair played the game iteratively, alternating their roles every round. The game was played in two consecutive sessions: first the players took sequential turns (turn-based), but in the second session they placed their tokens concurrently (concurrent). Conventional GLM analyses revealed activation throughout a diffuse collection of brain regions. In builders, during cooperation rounds we found bilateral activity in supplementary motor area, cerebellum and posterior, medial and anterior cingulate cortex, insula, caudate nucleus, putamen, hippocampus and amygdala. Insula, anterior and posterior cingulate cortex as well as amygdala have been repeatedly connected to different social cognitive processes, for example risk assessment (Bickart, Dickerson, & Feldman Barrett, 2014, Takahashi, Izuma, Matsumoto, Matsumoto, & Omori, 2015) or emotion communication (Anders, Heinze, Weiskopf, Ethofer, & Haynes, 2011). Supplementary motor area and frontal cortex activations may be

reflecting the planning of the next move. In the turn-based condition we can also see stronger activations of supplementary motor area that may reflect the same thing. This is possibly because, while one player is placing the token, the other has time to plan his next move. In the concurrent condition, player is placing his token and simultaneously is evaluating other player's moves, hence there is much less time to plan the next move. In competitive Turn-based trials builder showed activations in parietal cortex, insula and medial and anterior cingulate cortex bilaterally, left medial frontal gyrus, right superior frontal gyrus and right cerebellum. In Concurrent trials activations were present bilaterally in anterior and medial cingulate cortex. Anterior cingulate cortex has been widely connected to social cognition, decision making and empathy (Laura Astolfi et al., 2010; Schilbach et al., 2013) as well as conflict monitoring and cognitive control (Sebanz, Rebbechi, Knoblich, Prinz, & Frith, 2007). Helpers showed activations in precuneus and cuneus, medial cingulate cortex, right and left hippocampus and insula and also bilaterally in putamen. Precuneus activations are also reported in social cognition research, probably linked to self-other distinction processes (Fett, Shergill, & Krabbendam, 2015; Spiegelhalder et al., 2014), but precuneus has also been connected to Theory of mind processes (Carlson, Koenig, & Harms, 2013). Putamen activity was reported in cooperation tasks (Krill & Platek, 2012; Pfeiffer, Vogeley, & Schilbach, 2013). Hinderers showed activation in frontal lobes, parietal cortex, insula, thalamus and cerebellum when builders correctly placed a token in competition trials. Cerebellum is associated mostly with motor control, however there is more and more evidence suggesting that the role of cerebellum is much more diverse and that it may even take part in mirror network and mentalizing processes (Van Overwalle, Baetens, Mariën, & Vandekerckhove, 2014). Frontal and prefrontal activation most likely represent the attentional and executive demands of the task. The differences in turn-based and concurrent condition indicate less time for planning and bigger attentional load in concurrent condition. In inter-subject correlations, we found several clusters with correlated signal, mainly in superior frontal cortex for all conditions and in precuneus. This analysis, however, was only distinguishing between different conditions, not between more specific events. These results show an extensive pattern of activations in each of the conditions, with engaging many structures involved in social cognition. However, here we present only very simple analyses and further investigations are in progress, mainly focusing on inter-brain effects (e.g. generalised psychophysiological analyses- modelling activations in one player's brain per seed region in other player's brain), since we believe, that this is the biggest advantage that hyperscanning can provide.

Acknowledgements

This proceeding was supported by platform Czech-BioImaging within Czech-BioImaging Open access project. Stimulus materials were kindly provided by Tao Liu, Phd.

References

- Anders, S., Heinze, J., Weiskopf, N., Ethofer, T., & Haynes, J. D. (2011). Flow of affective information between communicating brains. *NeuroImage*, 54(1), 439–446.
<https://doi.org/10.1016/j.neuroimage.2010.07.004>
- Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., He, E. J., Roy, A., ... Babiloni, F. (2012). Cortical activity and functional hyperconnectivity by simultaneous EEG recordings from interacting couples of professional pilots. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4752–4755). IEEE. <https://doi.org/10.1109/EMBC.2012.6347029>
- Astolfi, L., Toppi, J., De Vico Fallani, F., Vecchiato, G., Salinari, S., Mattia, D., ... Babiloni, F. (2010). Neuroelectrical Hyperscanning Measures Simultaneous Brain Activity in Humans. *Brain Topography*, 23(3), 243–256.
<https://doi.org/10.1007/s10548-010-0147-9>
- Babiloni, C., Buffo, P., Vecchio, F., Marzano, N., Del Percio, C., Spada, D., ... Perani, D. (2012). Brains “in concert”: Frontal oscillatory alpha rhythms and empathy in professional musicians. *NeuroImage*, 60, 105–116.
<https://doi.org/10.1016/j.neuroimage.2011.12.008>
- Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J. P., ... Hari, R. (2012). MEG dual scanning: a procedure to study real-time auditory interaction between two persons. *Frontiers in Human Neuroscience*, 6(April), 1–7.
<https://doi.org/10.3389/fnhum.2012.00083>
- Beckmann, C. F. (2012). Modelling with independent components. *Neuroimage*, 62(2), 891–901.
- Bhaganagarapu, K., Jackson, G. D., & Abbott, D. F. (2013). An automated method for identifying artifact in independent component analysis of resting-state fMRI. *Frontiers in human neuroscience*, 7, 343.
- Bickart, K. C., Dickerson, B. C., & Feldman Barrett, L. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, 63, 235–248.
<https://doi.org/10.1016/j.neuropsychologia.2014.08.013>
- Billeke, P., Zamorano, F., López, T., Rodriguez, C., Cosmelli, D., & Aboitiz, F. (2014). Someone has to give in: theta oscillations correlate with adaptive behavior in social bargaining. *Social Cognitive and Affective Neuroscience*, 9(12), 2041–2048.
<https://doi.org/10.1093/scan/nsu012>
- Burgess, A. P. (2013). On the interpretation of synchronization in EEG hyperscanning studies: a cautionary note. *Frontiers in Human Neuroscience*, 7(December), 881.
<https://doi.org/10.3389/fnhum.2013.00881>
- Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(4), 391–402.
<https://doi.org/10.1002/wcs.1232>
- Decety, J., Jackson, P. L., Sommerville, J. a., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage*, 23(2), 744–751.
<https://doi.org/10.1016/j.neuroimage.2004.05.025>
- Fett, A. K. J., Shergill, S. S., & Krabbendam, L. (2015). Social neuroscience in psychiatry: unravelling the neural mechanisms of social dysfunction. *Psychological Medicine*, 45(6), 1145–65. <https://doi.org/10.1017/S0033291714002487>
- Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. *Neuron*, 88(1), 181–193.
<https://doi.org/10.1016/j.neuron.2015.09.022>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2), 782–790.
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(45), 16064–9.
<https://doi.org/10.1523/JNEUROSCI.2926-12.2012>
- Kasai, K., Fukuda, M., Yahata, N., Morita, K., & Fujii, N. (2015). The future of real-world neuroscience: Imaging techniques to assess active brains in social environments. *Neuroscience Research*, 90, 65–71.
<https://doi.org/10.1016/j.neures.2014.11.007>
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science (New York, N.Y.)*, 308(5718), 78–83.
<https://doi.org/10.1126/science.1108062>

- Konvalinka, I., & Roepstorff, A. (2012, January). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2012.00215>
- Krill, A. L., & Platek, S. M. (2012). Working together may be better: Activation of reward centers during a cooperative maze task. *PLoS ONE*, 7(2), 1–7. <https://doi.org/10.1371/journal.pone.0030613>
- Kuhl, J., & Kazén, M. (2002). PSSI-Inventář stylů osobnosti a poruch osobnosti. Praha: Testcentrum.
- Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neuroscience*, 10, 22. <https://doi.org/10.1186/1471-2202-10-22>
- Liu, T., & Pelowski, M. (2014). Clarifying the interaction types in two-person neuroscience research. *Frontiers in Human Neuroscience*, 8(April), 276. <https://doi.org/10.3389/fnhum.2014.00276>
- Ota, H., Fujii, T., Suzuki, K., Fukatsu, R., & Yamadori, A. (2001). Dissociation of body-centered and stimulus-centered representations in unilateral neglect. *Neurology*, 57(11), 2064–2069. <https://doi.org/10.1212/WNL.57.11.2064>
- Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience and Biobehavioral Reviews*, 37(10), 2516–2528. <https://doi.org/10.1016/j.neubiorev.2013.07.017>
- Sänger, J., Müller, V., & Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Frontiers in Human Neuroscience*, 7(June), 234. <https://doi.org/10.3389/fnhum.2013.00234>
- Sebanz, N., Rebbechi, D., Knoblich, G., Prinz, W., & Frith, C. D. (2007). Is it really my turn? An event-related fMRI study of task sharing. *Social Neuroscience*, 2(2), 81–95. <https://doi.org/10.1080/17470910701237989>
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4), 393–414. <https://doi.org/10.1017/S0140525X12000660>
- Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., Tebartz van Elst, L., Weiller, C., ... Tüscher, O. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behavioural Brain Research*, 258, 75–9. <https://doi.org/10.1016/j.bbr.2013.10.015>
- Takahashi, H., Izuma, K., Matsumoto, M., Matsumoto, K., & Omori, T. (2015). The anterior insula tracks behavioral entropy during an interpersonal competitive game. *PLoS ONE*, 10(6), 1–17. <https://doi.org/10.1371/journal.pone.0123329>
- Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *NeuroImage*, 86, 554–572. <https://doi.org/10.1016/j.neuroimage.2013.09.033>

Bariéry (nielen) psychologickej vedy: Kognitívne odchýlky vo vedeckom myslení

Jakub Šrol

Ústav experimentálnej psychológie
Centrum spoločenských a psychologických vied, Slovenská akadémia vied
Dúbravská cesta 9, Bratislava
jakub.srol@savba.sk

Abstrakt

Psychológovia identifikovali rad systematických odchýlok, ktoré ovplyvňujú ľudské myslenie v cele jeho šírke. Spomedzi mnohých možno uviesť napr. sklon k sebapotvrdzaniu, ktorý závažne ovplyvňuje schopnosť testovania hypotéz (Evans, 2016) a vyhľadávania a hodnotenia evidencie (Nickerson, 1998). Naše uvažovanie je taktiež skreslené v prospech dosahovania osobných cieľov, ktoré nemusia byť vždy v súlade so správnym rozhodovaním (Kunda, 1990). Cieľom príspevku bude upriamiť pozornosť na potenciálne skreslenia vo vedeckom myslení a to najmä s ohľadom na na liehavé problémy, s ktorými sa súčasná psychologická veda potýka, ako nízka replikabilita psychologického výskumu, či vysoká prevalencia využívania „sporných výskumných praktík“.

1 Úvod

Posledných pár rokov čeli psychológia ostrému návalu kritiky a pesimizmu. V prestížnych odborných časopisoch (najmä v časopisoch vydávaných Asociáciou pre psychologickú vedu, z angl. Association for Psychological Science) sa v súčasnosti venuje veľa priestoru témat, akými sú napríklad nízka replikabilita a malý dôraz na replikácie v psychologickom výskume (Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012), vysoká prevalencia využívania „sporných výskumných praktík“ (John, Loewenstein, & Prelec, 2012), či všeobecným pochybnostiam o adekvatnosti zaužívaných spôsobov vedenia výskumu, analýzy dát a publikácie získaných poznatkov (napr. Giner-Sorolla, 2012; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011; Wicherts et al., 2016). Niektoré z týchto problémov získali pozornosť aj širšej vedeckej komunity a v spojení s medializáciou nedávnych prípadov vedeckých podvodov v psychológií (Stroebe, Postmes, & Spears, 2012), prenikli tiež do povedomia širšej verejnosti. Môžeme teda hovoriť o naštrenení dôvery ľudí v psychologickú vedu¹.

Zatiaľ čo vedci, v ktorých záujme je vysoká kvalita výskumu, volajú po zmene zabehnutého systému, táto zmena nemôže nastaviť bez podrobnej analýzy príčin vyššie uvedených problémov. V tomto príspevku budem argumentovať, že viaceré z týchto problémov môžu byť dôsledkom kognitívnych skreslení, ktoré sú predmetom výskumu v kognitívnej a sociálnej psychológií (napr. Pohl, 2004). Uvediem príklady toho, ako niekoľko vybraných odchýlok vo vedeckom myslení môžu systematicky negatívne ovplyvňovať kvalitu zistení v psychológií. Nakoniec sa pokúsim načrtiť tiež niekoľko spôsobov, ktorými by bolo možné znížiť, alebo odstrániť negatívny vplyv uvedených kognitívnych skreslení a smerovať tak k vyššej kvalite kumulatívnej psychologickej vedy.

2 Kognitívne odchýlky vo vedeckom myslení

Psychologickí výskumníci identifikovali hned niekoľko kognitívnych skreslení, ktoré ovplyvňujú to, akým spôsobom si vyhľadávame a interpretujeme informácie, hodnotíme dostupnú evidenciu, či vytvárame a testujeme hypotézy. V nasledujúcej časti príspevku popíšem niektoré z týchto skreslení a uvediem, akým spôsobom môžu ovplyvňovať proces vedeckej práce v jeho rôznych fázach.

2.1 Skreslené hodnotenie evidencie: zaujatost' môjho pohľadu

Ked' sa púšťame do výskumu akéhokoľvek problému, našou úlohou je najskôr preštudovať si literatúru o fenoméne nášho záujmu a zhodnotiť stav výskumu v danej oblasti. V tejto časti našej práce sa snažíme kriticky zhodnotiť práce ostatných výskumníkov (príp. aj svoje predchádzajúce práce v danej oblasti), odhaliť ich silné stránky a limity a na základe toho posúdiť, v akom smere by sa mohol uberať ďalší výskum. Hoci predstava dokonalého vedca môže naznačovať, že výskumník v tejto fáze nezaujato zhodnotí všetku

¹ Zatiaľ čo v článku sa venujem súčasnej psychológií, väčšina spomenutých problémov nie je spojených výhradne s ňou, ale naopak, vzťahuje sa i na rôzne iné oblasti vedeckého bádania,

napríklad medicínu, biológiu, neurovedy, ekonómiu, či sociológiu (Fanelli, 2010, 2012; Ioannidis et al., 2014; John et al., 2012; Smaldino & McElreath, 2016).

dostupnú evidenciu, na základe čoho si neskôr vyvodí závery pre svoj ďalší výskum, štúdie kognitívnych odchýlok ukazujú, že táto predstava je zrejme zjednodušená.

Konkrétnie, výskum zaujatosti môjho pohľadu (z angl. myside bias) ukazuje systematickú zaujatosť ľudí svojimi počiatočnými názormi pri generovaní a hodnotení evidence (Macpherson & Stanovich, 2007). Štúdie tejto kognitívnej odchýlky dokladajú, že pokial' máme na nejakú problematiku utvorený istý názor, sme schopnejší generovať viac argumentov v súlade s našim názorom ako argumentov, ktoré ho vyvracajú (Toplak & Stanovich, 2003). Ešte dôležitejšie je však to, že ked' máme zhodnotiť kvalitu evidencie relevantnej pre naše počiatočné postoje (napr. vo forme argumentov, alebo vedeckých výskumov), sme kriticejší k evidencii, ktorej závery sú v rozpore s našim stanoviskom, a naopak, v prípade evidencie, ktorá podporuje naše stanovisko, prehliadame aj jej vážnejšie nedostatky (Kłaczynski, 2000; Stanovich & West, 1997).

Nedávny výskum Bastardiho, Uhlmann a Rossa (2011) zároveň naznačuje, že ešte dôležitejšiu úlohu ako predchádzajúce presvedčenia môžu v hodnotení vedeckej evidencie zohrávať naše túžby, to, čo si prajeme, aby bola pravda. Pokial' sa ako výskumníci nejakej oblasti venujeme dlhšiu dobu, veľmi pravdepodobne už inklinujeme k nejakej teórii a máme tak vybudované isté názory na fenomény, o ktoré sa zaujímame. Taktiež sme pritom poháňaní túžbou overiť si platnosť našich teórií – nájsť empirickú evidenciu, ktorá potvrdzuje naše hypotézy, zostaviť experiment, ktorý vyvracia platnosť opozičnej teórie. Tak postopej ako aj túžby nás zrejme vedú k tomu, hodnotiť každý ďalší poznatok, na ktorý narazíme, spôsobom, aký upevňuje naše presvedčenia.

2.2 Zaujaté vyhľadávanie informácií a tendencia ku konfirmácii hypotéz: sklon k sebapotvrdzovaniu

Pasívnym hodnotením dostupnej evidencie v prospech našich presvedčení a túžob však úloha kognitívnych odchýlok vo vedeckom myslení ani zdľaveka nekončí. Jedným z najrozšírenejších skreslení, ktoré je bezpochyby relevantné pri diskusii o vedeckom myslení, je sklon k sebapotvrdzovaniu (z angl. confirmation bias). Jadrom tejto odchýlky je, že naše existujúce presvedčenia ovplyvňujú spôsob, akým vyhľadávame a interpretujeme akékolvek ďalšie informácie relevantné pre tieto presvedčenia (Nickerson, 1998). Na rozdiel od pasívnej úlohy, ktorú pri hodnotení evidencie zohráva zaujatosť môjho pohľadu, pri sklone k sebapotvrdzovaniu máme tiež tendenciu aktívne vyhľadávať a interpretovať poznatky a testovať hypotézy tak, aby potvrdili naše existujúce presvedčenia (Evans, 2016).

Sú prinajmenšom tri spôsoby, akými sklon k sebapotvrdzovaniu ovplyvňuje vedecké myslenie (Nickerson, 1998). Po prvej, naše presvedčenia a túžby ovplyvňujú to, aké informácie vyhľadávame. Po tom,

čo si vytvoríme vlastné stanovisko, vyhľadávame a preferujeme informácie, ktoré sú s ním v súlade a vyhýbame sa tým, ktoré mu odporučujú (Jonas, Schulz-Hardt, Frey, & Thelen, 2001). Ďalej, pokial' sme vystavení komplexnej evidencii, napríklad vo forme výskumných štúdií, máme sklon interpretovať si ju v takom smere, aký utvrdzuje naše predchádzajúce presvedčenia. V šokujúcom výskume Lorda, Rossa a Leppera (1979) sa ukázalo, že pokial' majú ľudia s opačnými názormi zhodnotiť rovnaký výskum, ktorého závery sú mnohoznačné, každá zo skupín si závery výskumu interpretuje v prospech svojho vlastného stanoviska. Názory ľudí sú po zhodnotení výskumu dokonca ešte viac polarizované, ako boli pred vystavením sa tejto komplexnej informácii, keďže obe skupiny výskum vnímajú ako evidenciu pre svoj existujúci postoj. Nakoniec, už rané výskumy Wasona (1960) ukazujú, že pri testovaní hypotéz sa spontánne snažíme o potvrdenie našich presvedčení, a nie o ich vyvrátenie. Táto stratégia, nazývaná tiež sklon k pozitívemu testovaniu je však v rozpore s tradičným, popperovským názorom na testovanie hypotéz, podľa ktorého najviac informácií získame práve snahou o falzifikáciu hypotéz a teórií (Evans, 2016).

Sklon k sebapotvrdzovaniu sa úzko viaže k viacerým z problémov súčasnej psychologickej vedy. Vedie totiž k praktike nazývanej HARK-ovanie (z angl. hypothesizing after the results are known), ktorú Kerr (1998) načrtol už skoro pred dvadsiatimi rokmi, a ktorá spočíva v tom, akým spôsobom výskumníci pracujú s hypotézami. Zatial' čo podľa hypoteticko-deduktívneho modelu vedy by si výskumník mal stanoviť hypotézu a až následne zostaviť výskum, ktorým ju otestuje, takýto „konfirmačný“ výskumný postup sa v súčasnej psychologickej vede skoro vôbec neuplatňuje (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

V rámci oveľa bežnejšieho „exploračného“ výskumu psychologickej vedeči často najskôr preskúmajú získané dátá a na ich základe vyvodzujú hypotézy, ktoré následne analyzujú konfirmačnými štatistickými metódami. Sklon k sebapotvrdzovaniu nás v takejto situácii ľahko navedie na to, aby sme, ak objavíme v dátach čo i len náznak zaujímavej hypotézy, skúšali rôzne formy štatistickej analýzy a manipulácie s dátami až dovtedy, kým sa nám túto hypotézu nepodarí potvrdiť (Wagenmakers et al., 2012). Takýto prístup je navyše uľahčený ďalšou známyou kognitívnu odchýlkou – efektom spätného pohľadu (napr. Roese & Vohs, 2012), podľa ktorého preceňujeme predpovedateľnosť istej udalosti po tom, čo poznáme jej výsledok. Vďaka nej podľa Kerra (1998) môžeme získať pocit, ako keby „sme to celé už dopredu vedeli“ a dokonca môžeme uveriť, že hypotézy, ktoré sme vyvodili na základe získaných dát, boli v skutočnosti presne tým, o čo sme sa od začiatku výskumu najviac zaujimali.

Vyvodzovanie hypotéz z dát má mnoho teoretických a praktických negatívnych dopadov na kvalitu získaných výsledkov. V súčasnosti sa najviac diskutuje o tom, ako táto praktika môže pri analýze dát viest'

k veľkému nárustu štatistickej chyby typu II a následne zvyšenej proliferácií falošne pozitívnych výsledkov v psychológiu (Wagenmakers et al., 2012). HARK-ovanie má však aj menej zjavné následky, ktoré ale môžu mať z dlhodobého hľadiska ešte negatívnejší dopad na psychologickú vedu. Podľa Kerra (1998) sú pre výskumníkov silnými incentívami pre HARK-ovanie vysoké nároky na vedecké publikácie. Nakoľko býva výskum s jednoznačnými závermi podaný vo forme jednoduchého príbehu, v ktorom si stanovíme hypotézu a naše výsledky ju potvrdia, má vyššiu šancu na publikáciu. Nie je potom prekvapením, keď výskumníci využívajú HARK-ovanie na to, aby sa takémuto ideálnemu článku priblížili, a zvýšili si tak svoje šance na úspešné publikovanie.

Podobný argument nedávno rozvinul Giner-Sorolla (2012), ktorý hovorí o tom, ako zvýšené nároky a zároveň znížené možnosti publikovania v dôsledku pribúdajúceho množstva vedcov, snažiacich sa publikovať v obmedzenom množstve vedeckých časopisov, vedú výskumníkov k tomu, aby prezentovali perfektne vyzerajúce štúdie, ktoré väčšinou neodrážajú realitu vedeckej práce. Autor rozoberá, ako sa nároky na prezentáciu perfektne vyzerajúcich nových výsledkov jednoduchou a dobre čitateľnou formou v súčasnosti dostávajú do konfliktu so základným pilierom vedy – hľadaním pravdy. Skutočne, ľahšie spracovateľným podnetom už intuitívne silnejšie dôverujeme, páčia sa nám viac, a dokonca sme hlbšie presvedčení o ich pravdivosti (Alter & Oppenheimer, 2009). Okrem toho, výskumy tiež ukazujú, že ľudia hodnotia jednostranné argumenty ako kvalitnejšie v porovnaní s argumentmi, zohľadňujúcimi aj druhú stranu (Baron, 1995; Stanovich & West, 2008a).

Jednoducho a jednostranne podané publikované výstupy výskumov však veľmi pravdepodobne neodrážajú to, ako výskum naozaj prebiehal. Ako uvádzajú Giner-Sorolla (2012), tlak na podávanie informácií týmto spôsobom viedie napr. k tendencii publikovať iba štatisticky signifikantné výstupy. Tak posudzovatelia, ako aj samotní výskumníci sú silno zaujatí voči negatívnym výsledkom, ktoré následne tvoria minimálnu časť výstupov publikovaných v psychológii (Ioannidis, Munafo, Fusar-Poli, Nosek, & David, 2014). Táto averzia vytvára obrovské publikačné skreslenie, ktoré viedie k výraznému nadhodnoteniu efektov pozorovaných v psychológii, čo následne skresľuje meta-analytické odhady týchto efektov a znižuje pravdepodobnosť ich replikácie (Ioannidis et al., 2014; Open Science Collaboration, 2015).

2.3 Skreslené vnímanie pochybných výskumných praktík: motivované vedecké myslenie

V situácii, kedy sme v záujme dosiahnutia publikácie motivovaní skôr napĺňaním vyššie spomínaných „estetických štandardov“ (Giner-Sorolla, 2012), ako snahou o získanie pravdivých a kvalitných poznatkov, sa otvára priestor pre ďalšie kognitívne skreslenie, ktoré čiastočne vysvetľuje zlý stav súčasnej vedeckej

psychológie. Konkrétnie, vďaka mechanizmom motivovaného myslenia (z angl. motivated reasoning; Kunda, 1990) sme náhľadní zastávať postoje a využívať praktiky, ktoré nám najlepšie umožnia dospiť k želaným cieľom. V kompetitívnom prostredí súčasnej vedeckej psychológie, kde je našim cieľom získať čo najväčšie množstvo publikácií (Nosek, Spies, & Motyl, 2012), si môžeme prostredníctvom motivovaného myslenia ospravedlňovať postupy, ktoré zvyšujú publikovateľnosť na úkor prezentácie pravdivých, či najspoľahlivejších výsledkov. Takéto postupy, nazývané aj „sporné výskumné praktiky“² (z angl. questionable research practices), alebo „výskumníkove stupne voľnosti“ (z angl. researcher degrees of freedom) sú v súčasnosti predmetom intenzívnych debát vo vedeckých kruhoch (John et al., 2012; Simmons, Nelson, & Simonsohn, 2011).

V krátkosti, výskumník musí od naplánovania výskumu až po jeho publikáciu urobiť celý rad rozhodnutí týkajúcich sa vytvárania hypotéz a plánovania výskumu, zberu a analýzy dát, a publikovania výsledkov. Príkladom sú výber počtu participantov, premenných a štatistických analýz, či rozhodnutie o tom, ktoré výsledky budeme uvádzať v publikovanom výstupe z výskumu (Wicherts et al., 2016). Tieto rozhodnutia vytvárajú pre výskumníka veľký priestor na výber tých postupov, ktoré mu umožnia najvyššiu pravdepodobnosť dosiahnutia publikácie.

V tomto bode je dôležité povedať, že sporné výskumné praktiky väčšinou nie sú doménou podvodníkov – ich používanie je za istých situácií opodstatnené, vďaka čomu je pre výskumníkov ľahké si ich pomocou mechanizmov motivovaného myslenia ospravedlniť (Kunda, 1990) a následne je aj ich využívanie veľmi rozšírené. Vo výskume Johna a kol. (2012) sa ukázalo, že niektoré z týchto praktík psychológovia využívajú tak často, až sa stávajú de facto výskumnou normou.

Vyvstávajúcim problémom je, že využívanie sporných výskumných praktík môže viesť k enormnému nárustu pravdepodobnosti získania falošných pozitívnych výsledkov – ako napr. uvádzajú Simmons a kol. (2011): pokiaľ výskumník využije už iba štyri z najpoužívajúcich praktík, šanca na dosiahnutie signifikantného výsledku pri neprítomnosti pravého efektu sa môže zvýšiť až na 60%. Toto alarmujúce číslo, ktoré však stále môže byť ešte pomerne konzervatívnym odhadom (pozri Simmons et al., 2011), nás núti seriózne zvážiť dôveryhodnosť výsledkov publikovaných v odborných psychologickej kruhoch. Využívanie sporných výskumných

² Spomedzi mnohých sporných výskumných praktík môžem uviesť napríklad: priebežné rozhodovanie o veľkosti vzorky na základe toho, či nami predpokladané efekty dosiahli hladinu signifikancie; rozhodnutie neuvádzať v publikovanom výstupe všetky závislé premenné, alebo všetky vykonané štúdie, ale iba tie, ktoré vysli podľa našich očakávaní; prezentovanie neočakávaného výsledku v publikovanom výstupe takým spôsobom, ako keby sme ho od začiatku výskumu predpokladali (John et al., 2012)

praktík takisto prispieva k publikačnému skresleniu a následne aj k vyššie uvedeným problémom z neho vychádzajúcich (Wicherts et al., 2016).

3 Aké sú možnosti nápravy?

Uvedené kognitívne odchýlky ilustrujú, ako veľmi je proces vedeckej práce náhylný na zlyhania v rôznych fázach. Zatiaľ čo sa pomerne málo z výskumov skreslení orientovalo špecificky na skreslenia u vedeckých pracovníkov, neexistuje dôvod predpokladať, že tieto odchýlky sa u nich nevyskytujú v podobnej mieri, ako u zvyšku populácie.

Ako uvádzajú Ioannidis a kol. (2014), jedným z dôvodov prečo uvedené kognitívne odchýlky vo vedeckom myslení pretrvávajú napriek problémom, ktoré sú s nimi spojené je, že sami vedci nie sú schopní uvedomiť si svoje kognitívne odchýlky a kontrolovať ich vplyv. Existuje napríklad mnoho evidencie o tom, že pred kognitívnymi odchýlkami nás neochránia ani vyšše kognitívne schopnosti, resp. že ich vplyv je v tomto prípade relatívne slabý (Stanovich & West, 2008b). Konkrétnie, všeobecná inteligencia nemá vzťah k schopnosti odolať zaujatosti môjho pohľadu, či preferencii jednostranných argumentov (Klaczynski, 2000; Macpherson & Stanovich, 2007; Stanovich & West, 2008a).

Navýše, existujúca literatúra naznačuje, že pokial' aj máme poznatky o kognitívnych odchýlkach, stále považujeme samých seba za oveľa menej náhylných voči ich potenciálnemu vplyvu v porovnaní s inými ľuďmi – čo samo o sebe predstavuje ďalšie skreslenie nazývané aj efekt slepej škvarky (napr. West et al., 2012). Pokial' napokon aj spoľahlivo rozpoznáme, že nás intuitívne odchýlky vedú k nesprávnym záverom, napriek tomu často nie sme schopní ich vplyvu odolať (De Neys, 2013). Celkovo, tieto údaje naznačujú, že tréningy, vzdelenie, či vyšše kognitívne schopnosti nám nedokážu spoľahlivo zaručiť ochranu pred vplyvmi rôznych kognitívnych skreslení. Pri pokuse o ich nápravu je preto nutné hľadať iné riešenia.

Tieto možno hľadať v odstraňovaní zvyšných dvoch príčin pretrvávania kognitívnych odchýlok vo vedeckom myslení, ktoré uvádzajú Ioannidis a kol. (2014): to, že odchýlky ústiace do využívania sporných praktík si výskumníci ospravedlňujú pomocou mechanizmov motivovaného myslenia, a to, že vedecký systém je nastavený tak, že odmenuje publikovateľné poznatky na úkor pravdivých, čím implicitne odmenuje prejavu kognitívnych odchýlok, ktoré zvyšujú publikovateľnosť výstupov. Výskumníci navrhli hned' niekoľko spôsobov, ktoré by dokázali znížiť ospravedlniteľnosť sporných výskumných praktík prostredníctvom motivovaného myslenia a upraviť štruktúru motívacií tak, aby cieľom výskumníkov nebola publikácia ako taká, ale priamo akumulácia spoľahlivých vedeckých poznatkov (Giner-Sorolla, 2012; Ioannidis et al., 2014; Nosek et al., 2012; Simmons et al., 2011; Smaldino &

McElreath, 2016; Wagenmakers et al., 2012). Niektoré z nich stručne uvádzam nižšie.

3.1 Predregistrácia výskumov

Efektívny spôsob ako sa vysporiadať s HARK-ovaním podľa Wagenmakersa a kol. (2012) predstavuje predregistrácia výskumov. Pokial' si autori výskumu formulujú hypotézy a určia si presný výskumný postup vrátane spôsobu analýzy dát už pred samotným zberom dát, ich možnosť formulovať hypotézy ex post a prispôsobovať štatistické testy získaným dátam sa značne obmedzí, či úplne vytratí. Dostatočne špecifická a detailná predregistrácia by dokázala zabrániť aj využívaniu väčšiny sporných výskumných praktík (Wicherts et al., 2016).

3.2 Pravidlá pre výskumníkov a recenzentov

Ďalšiu možnosť navrhuje Simmons a kol. (2011), ktorí vytvorili niekoľko jednoduchých pravidiel pre výskumníkov a recenzentov. Ich dodržiavanie a kontrolovanie by zamedzilo zneužívaniu sporných výskumných praktík. Ako autori uvádzajú, sporné výskumné praktiky sú za istých situácií opodstatnené, preto je ich jednoduchšie si ich pred sebou obhájiť, a preto ich aj veľa používame. Treba preto zaviesť pravidlá, ktoré zmenia ich obhájiteľnosť, a jasne vymedzia, kedy sa tieto postupy smú a nesmú využívať.

Skutočne, výskum motivovaného myslenia naznačuje (Kunda, 1990), že pokial' odstráime výskumníkom možnosť obhájiť si pred sebou samým použitie sporných praktík, veľmi tým obmedzíme mechanizmy motivovaného myslenia, ktoré vedú k ich rozšírenému využívaniu. Zvýšenie transparentnosti výskumu napr. uvádzaním všetkých meraných premenných, alebo všetkých experimentálnych manipulácií (aj tých, ktoré zlyhalí) zníži obhájiteľnosť aspoň časti neadekvátnych postupov, ktoré výskumníci môžu využívať s cieľom dosiahnuť signifikantné výsledky (Simmons et al., 2011).

3.3 Priame replikácie

Zatiaľ čo požiadavka na publikovanie nových a prekvapivých výsledkov vytvára negatívny postoj k priamym replikáiam predchádzajúcich výskumov (Giner-Sorolla, 2012), priama replikácia predstavuje veľmi silný nástroj na zníženie proliferácie falóne pozitívnych výsledkov, zvýšenie reliability pozorovaných efektov a zníženie vplyvu publikačného skreslenia (Ioannidis et al., 2014; Open Science Collaboration, 2015). V tomto bode treba uviesť, že hoci sa v psychológií vyskytujú častejšie replikácie konceptuálne ako priame, a ich prínos pre zovšeobecnenie pozorovaných efektov je nespochybniatelný, konceptuálne replikácie nedokážu poslúžiť na zamedzenie vplyvu väčšiny problémov, ktorým som sa v tomto príspevku venoval (Nosek et al., 2012; Simmons et al., 2011).

3.4 Zmena motivačných štruktúr

Zatial' čo každý z vyššie uvedených spôsobov môže viesť k čiastočnému zlepšeniu problematického stavu súčasnej psychologickej vedy, skutočný posun správnym smerom bude zrejme vyžadovať zmenu systému, v ktorom ako vedci fungujeme. Ako uvádzajú Smaldino a McElreath (2016), mnohé z pochybných výskumných praktík sa mohli rozvinúť a ďalej sa šíria práve preto, že ich využívanie je vedeckým systémom odmeňované prostredníctvom zvýšenej pravdepodobnosti publikácie.

Podobný argument uviedol Kerr (1998) ako hlavný dôvod pre potenciálne vysokú proliferáciu HARK-ovania v psychologickom výskume. Podobne aj Nosek a kol. (2012) argumentujú, že pokým budú vo vedeckej psychológií incentívy pre získanie publikovateľných výstupov silnejšie ako tie pre získanie presných a spoľahlivých výstupov, vyššie uvedené problémy budú pretrvávať. Preto aj keď rôzne možné spôsoby kontroly výskumníkov majú šancu zamedziť využívaniu HARK-ovania a sporných výskumných praktík, oveľa praktickejším riešením je odstrániť motiváciu tieto praktiky využívať a nastaviť systém tak, aby výskumníci boli odmeňovaní za kvalitu svojich výstupov, nie za ich kvantitu.

Samozrejme, toto bude vyžadovať dlhodobú systematickú zmenu na inštitucionálnej úrovni, vedúcu k podporeniu praktík transparentného výskumu, otvorenému zdieľaniu dát a materiálov, a v neposlednom rade tiež atmosféry tolerancie voči nedokonalým výsledkom a učenia založenom na vlastných chybách (Giner-Sorolla, 2012; Ioannidis et al., 2014; Nosek et al., 2012; Smaldino & McElreath, 2016).

4 Záver

Súčasná psychologická veda je poznačená mnohými problémami, ktorým sa v odborných kruhoch venuje čoraz viac pozornosti. V tomto príspevku som sa pokúsil niektoré z týchto problémov – využívanie sporných výskumných praktík, HARK-ovanie, či publikačné skreslenie – vysvetliť ako následok podliehania niekoľkým vybraným odchýlkam v procese vedeckého myslenia. Kognitívne skreslenia sú notoricky ľažké na odstránenie prostredníctvom navodzovania intrapersonálnych zmien – tréningom, edukáciou, či rozvojom špeciálnych schopností. Ako potenciálne riešenia na odstránenie spomínaných problémov sa preto javí predregistrácia výskumov, nastolenie nových pravidiel pre výskumníkov a recenzentov a priame replikácie existujúcich výskumov. Zatial' čo tieto postupy majú potenciál zmieriť niektoré z uvedených problémov, jediným skutočným riešením by bola zrejme inštitucionálna zmena psychologickej vedy v prospech odmeňovania kvalitných a spoľahlivých výstupov a podporovania otvorenej vedeckej komunity.

Poděkování

Tento príspevok vznikol s podporou grantovej agentúry VEGA v rámci grantovej úlohy 2/0085/17 s názvom: Kognitívne limity efektívneho spracovania a komunikovania informácií.

Literatúra

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–35.
- Baron, J. (1995). Myside bias in thinking about abortion. *Thinking & Reasoning, 1*(3), 221–235.
- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful Thinking: Belief, Desire, and the Motivated Evaluation of Scientific Evidence. *Psychological Science, 22*(6), 731–732.
- De Neys, W. (2013). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning, 20*, 1–19.
- Evans, J. S. B. T. (2016). *How to Be a Researcher: A Strategic Guide for Academic Success*. New York: Routledge.
- Fanelli, D. (2010). Do pressures to publish increase scientist's bias? An empirical support from US states data. *PLoS ONE, 5*(4): e10271.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891–904.
- Giner-Sorolla, R. (2012). Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science. *Perspectives on Psychological Science, 7*(6), 562–571.
- Ioannidis, J. P. A., Munafo, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: Detection, prevalence, and prevention. *Trends in Cognitive Sciences, 18*(5), 235–241.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science, 23*(5), 524–532.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure

- to information. *Journal of Personality and Social Psychology*, 80(4), 557–571.
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217.
- Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: a two-process approach to adolescent cognition. *Child Development*, 71(5), 1347–66.
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning and Individual Differences*, 17(2), 115–127.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, 7(6), 615–631.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716-aac4716.
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530.
- Pohl, R. F. (2004). *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. New York: Psychology Press.
- Roese, N. J., & Vohs, K. D. (2012). Hindsight Bias. *Perspectives on Psychological Science*, 7(5), 411–26.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366.
- Smaldino, P. E., & McElreath, R. (2016). The Natural Selection of Bad Science. *Royal Society Open Science*, 3(9), 160384.
- Stanovich, K. E., & West, R. F. (1997). Reasoning Independently of Prior Belief and Individual Differences in Actively Open-Minded Thinking. *Journal of Educational Psychology*, 89(2), 342–357.
- Stanovich, K. E., & West, R. F. (2008a). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, 14(2), 129–167.
- Stanovich, K. E., & West, R. F. (2008b). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94(4), 672–695.
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science*, 7(6), 670–688.
- Toplak, M. E., & Stanovich, K. E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, 17(7), 851–860.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–32.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. a. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 627–633.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3), 129–140.
- West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, 103(3), 506–19.
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., Van Aert, R. C. M., & Van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 7(November), 1–12.

Where is my forearm? Clustering of body parts from simultaneous tactile and linguistic input using sequential mapping

Karla Štěpánová^{1,2}, Matěj Hoffmann^{1,3}, Zdeněk Straka¹, Frederico B. Klein⁴, Angelo Cangelosi⁴, Michal Vavrečka²

⁽¹⁾ Center for Machine Perception, Department of Cybernetics,

Faculty of Electrical Engineering, Czech Technical University in Prague

⁽²⁾ Czech Institute of Informatics, Robotics, and Cybernetics, CTU in Prague

⁽³⁾ iCub Facility, Istituto Italiano di Tecnologia

⁽⁴⁾ School of Computing, Electronics and Mathematics, Plymouth University, Plymouth, UK

Abstract

Humans and animals are constantly exposed to a continuous stream of sensory information from different modalities. At the same time, they form more compressed representations like concepts or symbols. In species that use language, this process is further structured by this interaction, where a mapping between the sensorimotor concepts and linguistic elements needs to be established. There is evidence that children might be learning language by simply disambiguating potential meanings based on multiple exposures to utterances in different contexts (cross-situational learning). In existing models, the mapping between modalities is usually found in a single step by directly using frequencies of referent and meaning co-occurrences. In this paper, we present an extension of this one-step mapping and introduce a newly proposed sequential mapping algorithm together with a publicly available Matlab implementation. For demonstration, we have chosen a less typical scenario: instead of learning to associate objects with their names, we focus on body representations. A humanoid robot is receiving tactile stimulations on its body, while at the same time listening to utterances of the body part names (e.g., hand, forearm and torso). With the goal at arriving at the correct “body categories”, we demonstrate how a sequential mapping algorithm outperforms one-step mapping. In addition, the effect of data set size and noise in the linguistic input are studied.

1 Introduction

Body representation has been the topic of psychological, neuroanatomical and neurophysiological studies for many decades. Spurred by the account of Head and Holmes (1911) and their proposal of superficial and postural schema, a number of different concepts has been proposed since: body schema, body image, and corporeal schema being only some of them. Body schema is usually thought of as more “low-level”, sensorimotor representation of the body used for action. Body image

is an umbrella term uniting higher level representations, for perception more than for action, and accessible to consciousness. Schwoebel and Coslett (2005) amassed evidence for distinguishing between three types of body representations: body schema, body structural description, and body semantics—constituting a kind of hierarchy. The body structural description is a topological map of locations derived primarily from visual input that defines body part boundaries and proximity relationships. Finally, body semantics is a lexical–semantic representation of the body including body part names, functions, and relations with artifacts (e.g., shoes are used on the feet, and feet can be used to kick a football).

While the details of every particular taxonomy or hierarchy can be discussed, clearly, there is a trend from continuous, modality-specific representations (like the tactile homunculus) to multimodal, more aggregated representations. This may be first instantiated by increasing receptive field size and combining sensory modalities, as it is apparent in somatosensory processing, e.g. areas relatively specialized on proprioception or touch and with small receptive fields (like Brodmann areas 3a and 3b), touch and proprioception are getting increasingly combined in areas 1 and 2. Then, going from anterior to posterior parietal cortex, the receptive fields grow further and somatosensory information is combined with visual. One can then ask whether this process of bottom-up integration or aggregation may give rise to discrete entities, or categories, similar to individual body parts. Vignemont et al. (2009) focused on how body segmentation between hand and arm could appear based on a combined tactile and visual perception. They explored category boundary effect which appeared when two tactile stimuli were presented: these stimuli felt farther away when they were applied across the wrist than when they were applied within a single body part (palm or forearm). In conclusion, they suggest that the representation of the body is structured in categorical body parts delineated by joints, and that this categorical representation modulates tactile spatial perception.

Next to the essentially bottom-up clustering of

multimodal body-related information, an additional “categorization” of body parts is imposed through language, such as when the infant hears her parents naming the body parts. Interestingly, recent research (Majid, 2010) showed that there are some cross-linguistic variabilities in naming body parts and this may in turn override or influence the “bottom-up” multimodal (non-linguistic) body part categorization.

While the field is relatively rich in experimental observations, the mechanisms behind the development and operation of these representations are still not well understood. Here, computational and in particular robotic modeling ties in—see (Hoffmann et al., 2010; Schillaci et al., 2016) for surveys on body schema in robots. Petit and Demiris (2016) developed an algorithm for the iCub humanoid robot to associate labels for body parts and later proto-actions with their embodied counterparts. These could then be recombined in a hierarchical fashion (e.g., “close hand” consists of folding individual fingers). Mimura et al. (2017) used Dirichlet process Gaussian mixture model with latent joint to provide a Bayesian body schema estimation based on tactile information. Their results suggest that kinematic structure could be estimated directly from tactile information provided by a moving fetus without any additional visual information—albeit with a lower accuracy. Our own work on the iCub humanoid robot has thus far focused on learning primary representations—tactile (Hoffmann et al., 2017) and proprioceptive (Hoffmann and Bednarova, 2016). In this work, we use the former (the “tactile homunculus”) as input for further processing—interaction with linguistic input.

In this work, we strive to find segmentation of body parts based on a simultaneous tactile and linguistic information. However, body part categorization and mapping to body part names is one instance of a more general problem of segmenting objects from the environment, learning compressed representations (loosely speaking: concepts, categories, symbols) to stand in for them and associating them with words to which the infant is often exposed simultaneously. Borghi et al. (2004), for example, studied the interaction of object names with situated action on the same objects.

We made use of a newly proposed sequential mapping algorithm which extends an idea of one-step mapping (Smith et al., 2006) and compared its overall accuracy to one-step mapping as well as to accuracies of segmenting individual body parts. We further explore how the accuracy of the learned mapping is influenced by a level of noise in the linguistic domain and data set size. The sequential mapping strategy was shown to be very robust as it can find the mapping under circumstances of very noisy input and clearly outperformed the one-step mapping.

Complete source code used for generating results in this article is publicly available at <https://github.com/stepakar/sequential-mapping>.

<https://github.com/stepakar/sequential-mapping>.

This article is structured as follows. The inputs and their preprocessing and the mapping algorithms are described in Section 2. This is followed by Results (Section 3) and a Discussion and Conclusion.

2 Materials and Methods

In this section, we will first present the inputs and their preprocessing pipelines: tactile input (Section 2.1) and linguistic input (Section 2.2). In total, 9 body parts of the right half of the robot’s upper body were stimulated: torso/chest, upper arm, forearm, palm and 5 fingertips. Tactile stimulation coincided with an utterance of the body part’s name. Then, the one-step and sequential mapping algorithms (sections 2.3.1 and 2.4) are presented, and a description of the evaluation (Section 2.5).

2.1 Tactile inputs and processing

To generate tactile stimulation pertaining to different body parts, we built on our previous work on the iCub humanoid robot. In particular, the “tactile homunculus” (Hoffmann et al., 2017)—a primary representation of the artificial sensitive skin the robot is covered with (see Fig. 1 – one half of the robot’s upper body). In the current work, the skin was not physically stimulated anymore, but the activations were emulated and then relayed to the “homunculus”, as detailed below.

2.1.1 Emulated tactile input

We created a YARP (Metta et al., 2006) software module to generate virtual skin contacts¹. A skin part was randomly selected and then stimulated. The number of pressure-sensitive elements (henceforth taxels) for different skin parts was 440 for the torso, 380 for upper arm, 230 for forearm, and 104 for the hand (44 for palm and 5×12 for fingertips)—1154 taxels in total. Once the skin part was randomly selected, a small region was also randomly picked within that part for the tactile stimulation—10 taxels at a time, corresponding to the triangular modules the skin is composed of. For the hand, the situation was slightly different: the entire hand was treated as one skin part. Then, within the hand, a random choice was made between 5 subregions on the palm skin (8 to 10 taxels) and 5 fingertips (12 taxels each). Data was collected for 100 minutes, corresponding to approximately 2000 individual 3 second stimulations. For all skin parts, the stimulation lasted for 3 seconds and was sampled at 10 Hz. A label–body part name—was saved along with the tactile data. These labels are used to generate the linguistic input and for performance evaluation later, but do not directly take

¹<https://github.com/robotology/peripersonal-space/tree/master/modules/virtualContactGeneration>

part in the clustering of tactile information. Please note that there were separate labels for the palm and individual fingers, while these were all treated as one “skin part” in the virtual touch generation and hence the number of samples per finger, for example, was lower than for other non-hand body parts.

2.1.2 First layer – “tactile homunculus”

The input layer of the “tactile homunculus” (Hoffmann et al., 2017) consists of a vector, $\mathbf{a}(t)$, of activations of 1154 taxels at time t —the output of the previous section—that have binary values (1 when a taxel is stimulated, 0 otherwise). The output layer then forms a 7×24 (168 “neurons” in total) grid – see Figure 1 B. This layer is a compressed representation of the skin surface—the receptive fields of neurons (the parts of skin they respond to) are schematically color-coded. However, this code (and “clustering”) is not available as part of the tactile input.

The output layer will be represented as a single vector $\mathbf{x}(t) = [x_1(t), \dots, x_{168}(t)]$. The activations of the output neurons, $x_i(t)$, are calculated as dot products of the weight vector \mathbf{u}_i corresponding to the i -th output neuron and the tactile activation vector $\mathbf{a}(t)$ as follows:

$$x_i(t) = \mathbf{u}_i \cdot \mathbf{a}(t) \quad (1)$$

2.1.3 Second layer – GMM

The output of the first layer, vector $\mathbf{x}(t)$ (168 elements, continuous-valued) serves as input to the second tactile processing layer. This layer aims to cluster individual body parts and represent them as abstract models. Resulting models T_j are subsequently mapped in the multimodal layer to clusters found in the language layer.

To process the outputs from the first layer, we used a Gaussian mixture model (GMM), which is a convex mixture of D -dimensional Gaussian densities $l(\mathbf{x}|\theta_j)$. In this case, each tactile model T_j is described by a set of parameters θ_j . The posterior probabilities $p(\theta_j|\mathbf{x})$ are computed as follows:

$$p(\theta_j|\mathbf{x}) = \sum_{j=1}^J r_j^k l(\mathbf{x}|\theta_j), \quad (2)$$

$$l(\mathbf{x}|\theta_j) = \frac{1}{\sqrt{(2\pi)^D} \sqrt{|\mathbf{S}_j|}} \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{m}_j)^T (\mathbf{S}_j)^{-1} (\mathbf{x} - \mathbf{m}_j)\right], \quad (3)$$

where \mathbf{x} is a set of D -dimensional continuous-valued data vectors, r_j^k are the mixture weights, J is the number of tactile models, parameters θ_j are cluster centers \mathbf{m}_j and covariance matrices \mathbf{S}_j .

Mixture of Gaussians is trained by the EM algorithm (Dempster et al., 1977). Number of tactile models J is in this model preset based on the number of different linguistic labels. In future, we plan to

use an adaptive extension of GMM algorithm such as gmGMM (Štepánová and Vavrečka, 2016) to detect this number autonomously.

An output of this layer for each data point $\mathbf{x}(t)$ is the vector $\mathbf{y}(t)$ of J output parameters describing the data point (the likelihood that the data point belongs to each individual cluster in a mixture). This corresponds to the fuzzy memberships (distributed representation).

2.2 Linguistic inputs and processing

Tactile stimulation of a body part was accompanied with the corresponding utterance. In our case, where we have 9 separate body parts, these are ‘torso’, ‘upper arm’, ‘forearm’, ‘palm’, ‘little finger’, ‘ring finger’, ‘middle finger’, ‘index finger’ and ‘thumb’. Linguistic and tactile inputs are processed simultaneously.

We conducted experiments with spoken language input—one-word utterances pronounced by a non-native English speaker. To process this data, we made use of CMU Sphinx (an open-source flexible Markov model-based speech recognizer system) (Lamere et al., 2003) and achieved 100% accuracy of word recognition. The word-forms are extracted from the audio input and compared to prelearned language models by means of the log-scale scores $p(\mathbf{w}_i^n | L_i)$ of the audio matching. Based on these data, posterior probability can be computed.

However, in the current work, we employed a shortcut and used the labels (ground truth) directly. This allowed us to fully explore the effect of misclassification in linguistic subdomain to mapping accuracy. The noise to the language data was added subsequently and evenly to all classes (a given proportion of labels was randomly permuted).

2.3 Cross-situational learning

One possible way how to establish mapping between sensorimotor concepts and linguistic elements is to use frequencies of referent and meaning co-occurrences, that is, the ones with the highest co-occurrence are mapped together (Smith et al., 2006; Xu and Tenenbaum, 2007). This method is usually called cross-situational learning and supposes the availability of the ideal associative learner who can keep track and store all co-occurrences in all trials, internally memorizing and representing the word–object co-occurrence matrix of input. This allows the learner to subsequently choose the most strongly associated referent (Yu and Smith, 2012).

2.3.1 One-step mapping

The simplest one-step word-to-referent learning algorithm only accumulates word-referent pairs. This can be viewed as Hebbian learning: the connection between a

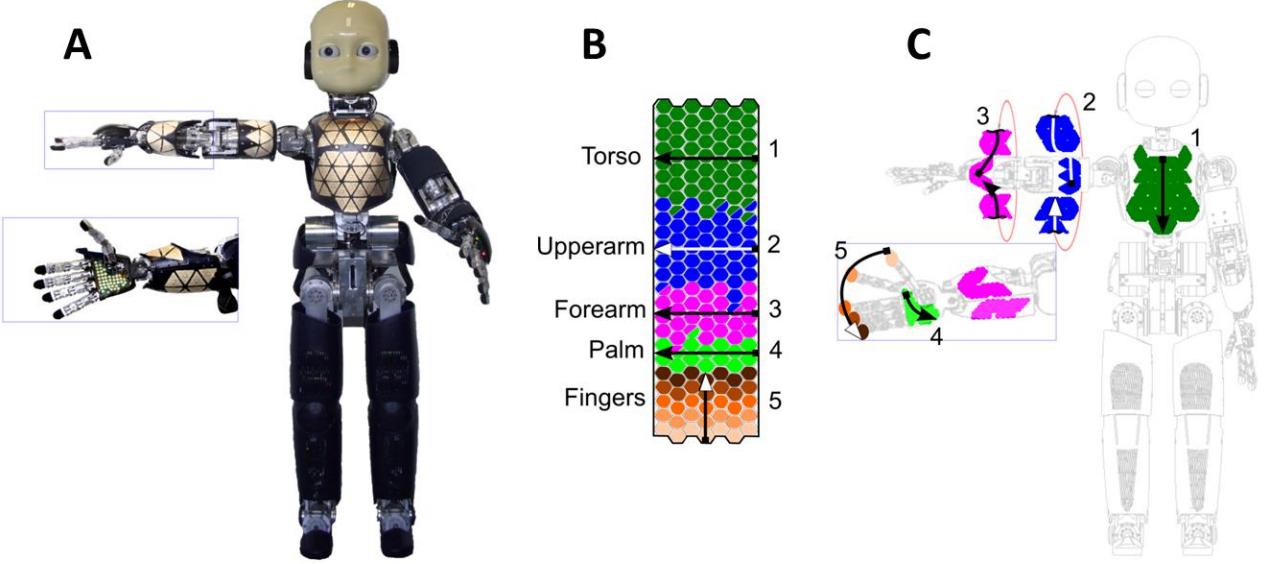


Fig. 1: iCub skin and tactile homunculus. (A) Photograph of the iCub robot with artificial skin exposed on the right half of the upper body (1154 taxels in total). (B) Representation of tactile inputs learned using a Self-Organizing Map – a 24×7 neuronal sheet. (C) Schematics with skin patches unfolded and colored to mark the correspondence with B. Arrows illustrate the relationship in orientation between skin parts and the learned map (Hoffmann et al., 2017).

word and an object is strengthened if the pair co-occurs in a trial. To extend this basic idea, we can enable also forgetting by introducing a parameter η , which can capture the memory decay (Yu and Smith, 2012). Supposing that at each trial t we observe an object o_t^n and hear a corresponding word w_t^n (N_t possible associations), we can describe the update of the strength of the association between word model $L(i)$ and object—in our case tactile model $T(j)$ —as follows:

$$A(i, j) = \sum_{t=1}^R \eta(t) \sum_{n=1}^{N_t} \delta(w_t^n, i) \delta(o_t^n, j), \quad (4)$$

where R is the number of trials, δ is the Kronecker delta function (equal to 1 when both arguments are identical and 0 otherwise), w_t^n and o_t^n indicate the n th word–object association that the model attends to and attempts to learn in the trial t and $\eta(t)$ is the parameter controlling the gain of the strength of association.

Now let's assume that the word $w(i)$ is modeled by the model L_i in the language domain and object (referent) $o(j)$ is modeled by the model $T_{m(i)}$ in the tactile domain. Our goal is to find the corresponding model $T_{m(i)}$ from tactile subdomain for each model L_i from language domain to assign them together. Indices $m(i)$ are found as follows:

$$\forall i : m(i) = \operatorname{argmax}_i A(i, j), \quad (5)$$

where \mathbf{A} is the co-occurrence matrix computed in the Eq. 4 (element $A(i, j)$ captures co-occurrence between the word $w(i)$ and object $o(j)$).

2.4 Sequential mapping algorithm

To capture dynamic competition among models, we extend the basic one-step mapping algorithm for cross-situational learning by sequential addition of inhibitory connections. The inhibitory mechanisms and situation-time dynamics were already partially included into the model of cross-situational learning proposed by McMurray et al. (2012). Even though our model shares some similarities with the model proposed by McMurray, it stems from different computational mechanisms. After a reliable assignment between a language and tactile model is found, inhibitory connections among this tactile model and all other language models are added. Thanks to this mechanism, mutual exclusivity principle (the fact that children prefer mapping where object has only one label to multiple labels (Markman, 1990)) is guaranteed.

The assignment between tactile models T_j and language models L_j is found using the following iterative procedure:

1. Tactile and language data are clustered separately and the corresponding posterior probabilities are found.
2. For each data point the most probable tactile and language clusters are selected and the data point is assigned to these clusters.
3. Co-occurrence matrix with elements $A(i, j)$ is computed and the best assignment is selected:

$$[im, m(im)] = \operatorname{argmax}_i \operatorname{argmax}_j A(i, j). \quad (6)$$

- In this step, the tactile model $T_{m(im)}$ is assigned to the language model L_{im} .
4. Inhibitory connections are added between the assigned tactile model $T_{m(im)}$ and all language models L_i , where $i \neq im$ (mutual exclusivity).
 5. Assigned data points (data points which belong to both $T_{m(im)}$ and L_{im}) are deleted from the data set.
 6. If data set is not empty or not all tactile clusters are assigned to some language cluster go to (1), else stop.

2.5 Evaluation

Accuracy of the learned mapping is calculated in the following manner: We cluster output activations from the tactile homunculus and assign each data point to the most probable cluster. Then, we find indices $m(i)$ for all clusters as defined in equation 5 for one-step mapping and equation 6 for sequential mapping. Based on this mapping we can assign each data point to the language label. These language labels are subsequently compared to the ground truth (the body part name is equivalent to the language label prior to the application of noise). Accuracy is then computed as:

$$acc = TP/N \quad (7)$$

where TP (true positive) is the number of correctly assigned data points and N is the number of all data points.

3 Results

We studied the performance of one-step vs. the sequential mapping algorithms on the ability to cluster individual body parts from simultaneous tactile and linguistic input. That is, all the skin regions on the same body part should “learn” that they belong together (to the forearm, say), thanks to the co-occurrences with the body part labels. In addition, the effect of data set size and levels of noise in the linguistic domain are investigated (Section 3.1). A detailed analysis of the mapping accuracy for individual body parts and a backward projection onto the tactile homunculus are shown in sections 3.2 and 3.3 respectively.

3.1 Comparison of accuracy of one-step mapping to sequential mapping

The performance of the one-step and sequential mapping algorithms is shown in Fig. 2. The comparison is provided for different data set sizes (namely for 6 different data sets with number of data points from 64 to 63806) and noise levels. As can be seen, the accuracy of sequential mapping remains very stable and outperforms one-step mapping for all values of the noise (in

the linguistic domain) and all data set sizes. For smaller data sets, we can see a steeper drop in accuracy with increasing noise in the language data.

3.2 Accuracy of mapping for individual body parts

The accuracy calculated in the previous section and Fig. 2 is an overall accuracy and we don’t take into account the number of data points per individual body part. To explore the performance in more detail, we focused also on the accuracy of sequential mapping for individual body parts. The results for the data set with 3190 and 638 data points can be seen in Fig. 3 top and bottom panel, respectively. The accuracy for all body parts decreases with increasing noise in the linguistic input. The accuracy for fingers is significantly lower—this is due to the lower number of samples per finger (see Section 2.1.1). Comparing the top and bottom panel in Fig. 3 demonstrates poorer performance with higher variance, especially for the fingers.

3.3 Projecting results of sequential mapping back onto homunculus

After tactile data from homunculus are clustered and these clusters are mapped to appropriate language clusters (representing body parts utterances), we can project these labels back onto the original tactile homunculus. Considering that $x_i(t)$ are activations of neuron i in the homunculus, D is the whole data set consisting of vector of homunculus activations for each data point, and $LangLabel(d)$ is the language label assigned to a data point d based on the sequential mapping procedure described in the Section 2.4, we can project results of sequential mapping onto the homunculus in a following manner. First, we compute strength of activation n_i^k of each neuron i for a given language label k as follows:

$$n_i^k = \sum_{x_i(t) \in D_k} x_i(t), \quad i \in \{1, \dots, 168\}, \quad (8)$$

where $D_k = \{d \in D | LangLabel(d) = k\}$ and $k = \{\text{torso, upper arm, forearm, palm, little finger, ring finger, middle finger, index finger, thumb}\}$.

Afterwards, we visualize for each neuron how much it is activated for individual body parts. Results for data sets of differing size and level of noise in the linguistic domain can be seen in Fig. 4. Clearly, for large enough data sets and limited noise, the mapping from language to the tactile modality is successful in delineating the body part categories (the fingers with fewer data points being more challenging)—as can be seen by comparing panels A and B.

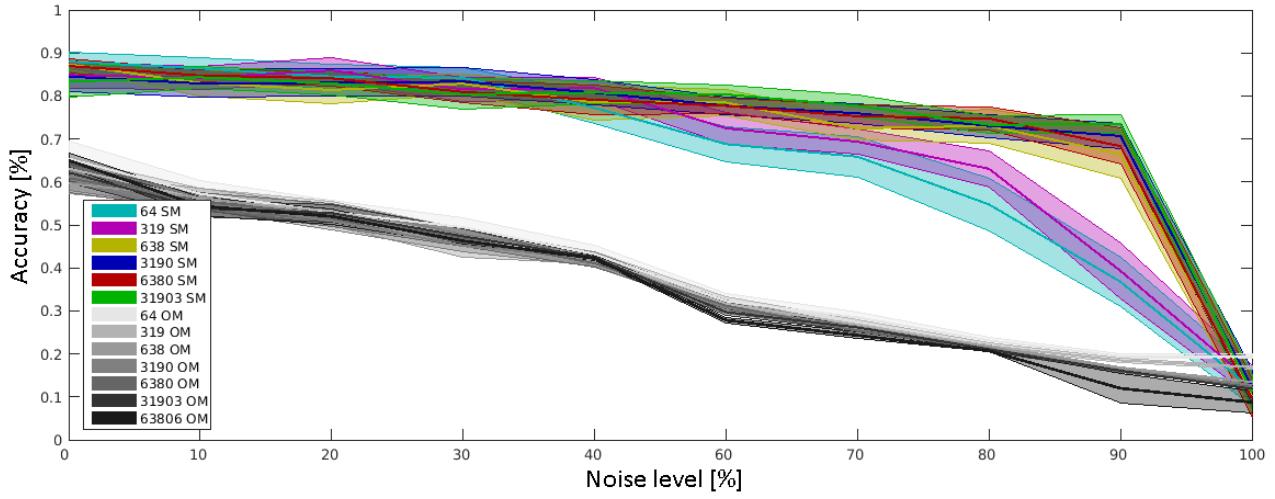


Fig. 2: Accuracy of one-step vs. sequential mapping for different levels of noise in language. Number denotes the size of data set, SM - sequential mapping and OM - one-step mapping. The mean and standard deviation from 20 repetitions is visualized.

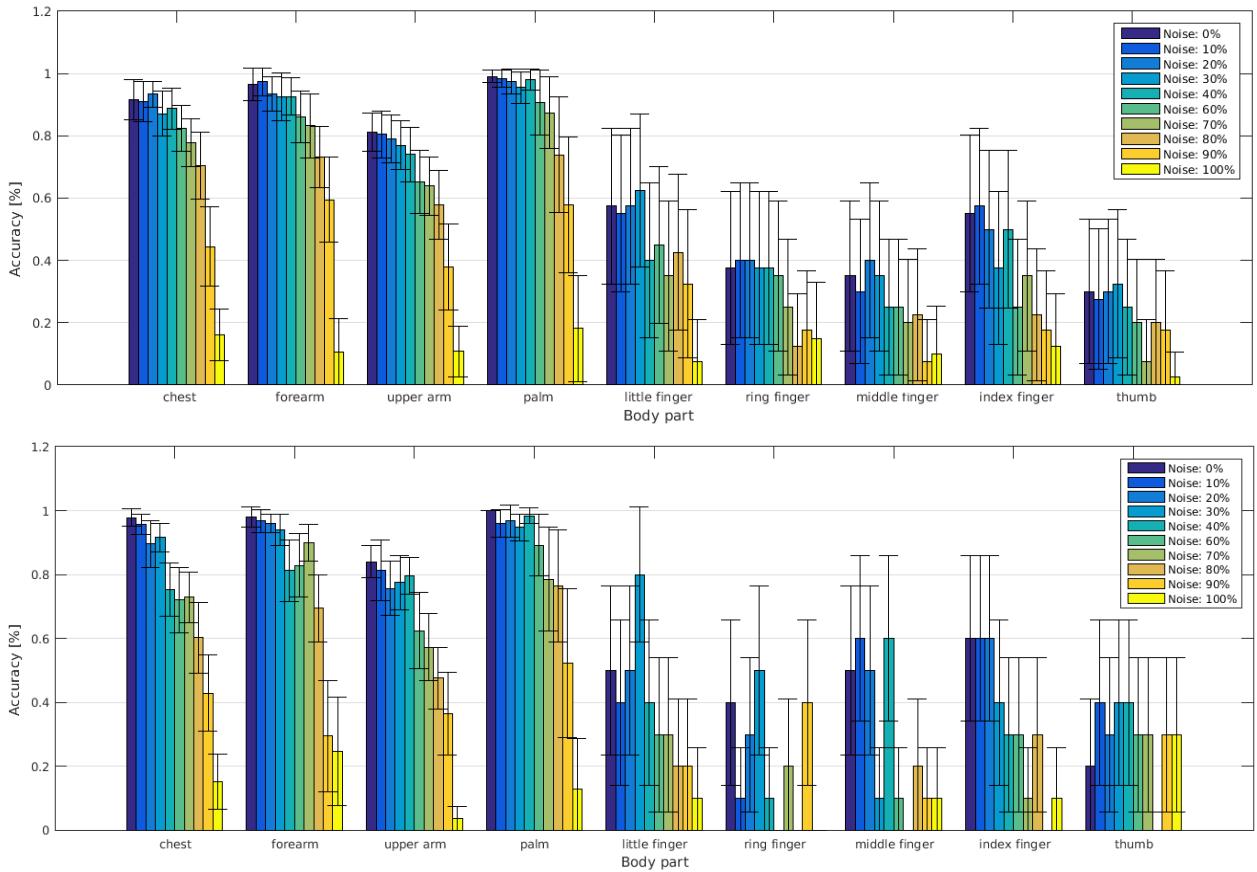


Fig. 3: Accuracy of sequential mapping for individual body parts: Visualization of sequential mapping accuracy based on the noise in linguistic data for 2 data set sizes: 3190 data points (upper) and 638 data points (lower), noise in language data 0-100%(random). The mean and standard deviation from 40 repetitions are visualized.

4 Discussion and Conclusion

To study the problem of associating (mapping) between sensorimotor or multimodal information, concepts or

categories, and language or symbols, we have chosen a specific but less studied instance of this problem: segmentation and labeling of body parts. Perhaps, from a developmental perspective, this could be plausible, as

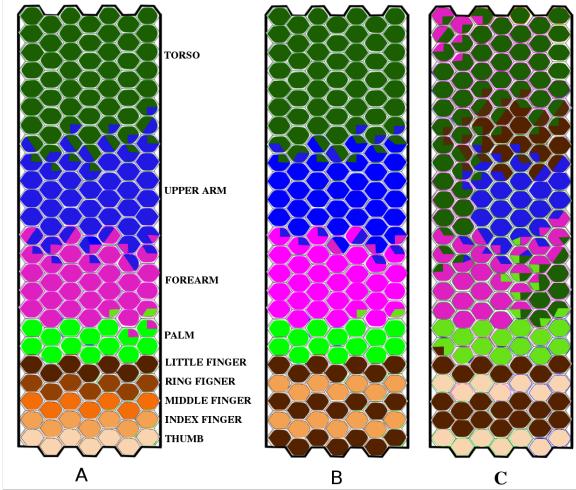


Fig. 4: Projection of mapping results back onto the tactile homunculus – sample runs of the algorithm. Color code for individual body parts is the same as in Fig. 1. (A) Original homunculus with true labels. (B) Results from data set with 6381 data points and 10% noise. (C) Results from 638 data points and 80% noise.

the body may be the first “object” the infant is discovering. The self-exploration occurs in the sensorimotor domain, but at the same time or slightly later, the infant is exposed to utterances of body part names. In this work, we study the mapping between the tactile modality and body part labels from linguistic input.

We present a new algorithm for mapping language to sensory modalities (sequential mapping), compare it to one-step mapping and test it on the body part categorization scenario. Our results suggest that this mapping procedure is robust, resistant against noise, and sequential mapping shows better performance than one-step mapping for all data set sizes and also slower performance degradation with increasing noise in the linguistic input. Furthermore, we explored accuracy of the sequential mapping for individual body parts, revealing that body parts less represented in the data set—fingers—were categorized less accurately. This problem might be mitigated with increased overall data set size; yet, dealing with clusters with uneven data point number is a common problem of clustering algorithms (in our case GMM).

Projecting the labels or categories induced by language back onto the tactile homunculus showed that the

body part categories are quite accurate. Given the nature of the tactile input—the skin is a continuous receptor surface—and the random-uniform tactile input generator used, the linguistic input was the only one that can facilitate cluster formation. However, more realistic, non-uniform touch and, in particular, the addition of additional modalities (proprioception, vision) should enable bottom-up non-linguistic body part category formation, as described by (Vignemont et al., 2009), for example. These constitute possible directions of our future work: the “modal” cluster formation will interact with the labels imposed by language. Furthermore, thus far, only one half of the body was considered—corresponding to the lateralized representations in the tactile homunculus—but one can imagine stimulating both left and right arm, for example, while hearing always the same utterance: ‘upper arm’. Further study of the brain areas involved in this processing is needed, in order to develop models more closely inspired by the functional cortical networks, like in (Caligiore et al., 2010) that model the experimental findings of (Borghi et al., 2004).

For our experiments we used artificially generated linguistic input (i.e., body part labels) with added noise (i.e. wrong labels with a certain probability). In the future, we are planning to use actual auditory input (spoken words) with real noise. This will also add the additional dimension of similarity in the auditory domain: ‘arm’ and ‘forearm’ are phonetically closer to each other than to, say, ‘torso’. Thus, the linguistic modality will not constitute crisp, discrete labels anymore, but these will have to be extracted first—opening up further possibilities for bidirectional interaction with other modalities.

5 Acknowledgement

K.S. and M.H. were supported by the Czech Science Foundation under Project GA17-15697Y. M.H. was additionally supported by a Marie Curie Intra European Fellowship (iCub Body Schema 625727) within the 7th European Community Framework Programme. Z.S. was supported by The Grant Agency of the CTU Prague project SGS16/161/OHK3/2T/13. M.V. was supported by European research project TRADR funded by the EU FP7 Programme, ICT: Cognitive systems, interaction, robotics (Project Nr. 609763).

References

- Borghi, A. M., Glenberg, A. M. and Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, 32(6):863–873.
- Caligiore, D., Borghi, A. M., Parisi, D. and Baldassarre, G. (2010). Tropicals: A computational embodied

- neuroscience model of compatibility effects. *Psychological Review*, 117(4):1188.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.
- Head, H. and Holmes, H. G. (1911). Sensory disturbances from cerebral lesions. *Brain*, 34:102–254.
- Hoffmann, M. and Bednarova, N. (2016). The encoding of proprioceptive inputs in the brain: knowns and unknowns from a robotic perspective. Vavrecka, M., Becev, O., Hoffmann, M. and Stepanova, K. (eds.), In *Kognice a umělý život XVI [Cognition and Artificial Life XVI]*, pp. 55–66.
- Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M. and Pfeifer, R. (2010). Body schema in robotics: A review. *Autonomous Mental Development, IEEE Transactions on*, 2(4):304–324.
- Hoffmann, M., Straka, Z., Farkas, I., Vavrecka, M. and Metta, G. (2017). Robotic homunculus: Learning of artificial skin representation in a humanoid robot motivated by primary somatosensory cortex. *IEEE Transactions on Cognitive and Developmental Systems*.
- Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., Warmuth, M. and Wolf, P. (2003). The cmu sphinx-4 speech recognition system. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong*, vol. 1, pp. 2–5. Citeseer.
- Majid, A. (2010). Words for parts of the body. *Words and the mind: How words capture human experience*, pp. 58–71.
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1):57–77.
- McMurray, B., Horst, J. S. and Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological review*, 119(4):831.
- Metta, G., Fitzpatrick, P. and Natale, L. (2006). Yarp: yet another robot platform. *International Journal on Advanced Robotics Systems*, 3(1):43–38.
- Mimura, T., Hagiwara, Y., Taniguchi, T. and Inamura, T. (2017). Bayesian body schema estimation using tactile information obtained through coordinated random movements. *Advanced Robotics*, 31(3):118–134.
- Petit, M. and Demiris, Y. (2016). Hierarchical action learning by instruction through interactive grounding of body parts and proto-actions. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 3375–3382. IEEE.
- Schillaci, G., Hafner, V. V. and Lara, B. (2016). Exploration behaviors, body representations, and simulation processes for the development of cognition in artificial agents. *Frontiers in Robotics and AI*, 3:39.
- Schwoebel, J. and Coslett, H. B. (2005). Evidence for multiple, distinct representations of the human body. *Journal of cognitive neuroscience*, 17(4):543–553.
- Smith, K., Smith, A. D., Blythe, R. A. and Vogt, P. (2006). Cross-situational learning: a mathematical approach. *Lecture Notes in Computer Science*, 4211:31–44.
- Štepánová, K. and Vavrečka, M. (2016). Estimating number of components in gaussian mixture model using combination of greedy and merging algorithm. *Pattern Analysis and Applications*, pp. 1–12.
- Vignemont, d. F., Majid, A., Jola, C. and Haggard, P. (2009). Segmenting the body into parts: evidence from biases in tactile perception. *The Quarterly Journal of Experimental Psychology*, 62(3):500–512.
- Xu, F. and Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological review*, 114(2):245.
- Yu, C. and Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological review*, 119(1):21.

Modelování života a organismus uměleckého díla

Aleš Svoboda

Univerzita Karlova, Fakulta humanitních studií
U Kříže 8, 158 00 Praha 5 – Jinonice
ales.svoboda@fhs.cuni.cz

Abstrakt

Příspěvek se na základě kritiky několika úspěšných uměleckých intervencí do simulací umělého života a nového promyšlení smyslu takových simulací ve vztahu k umění pokusí stanovit parametry návrhu organismu uměleckého díla. Podobně, jako biologický růst je zasazen do energetických a informačních poměrů ekosystému, je umělecké dílo rovnováhou jistých „energetických stavů“ a souborem jejich hierarchických vztahů.

1 Inteligence vidění

Margaret A. Boden, průkopnice kognitivní vědy, načrtla v úvodu knihy *AI, její povaha a budoucnost* (2016) charakter a smysl předmětu umělé inteligence poměrně jednoduše: „Umělá inteligence (AI) se snaží o to, aby počítače dělaly onen druh věcí, které dělají myslí. / Některé z nich (například uvažování) obvykle charakterizujeme jako „inteligentní“. Jiné zase (například „vidění“) nikoliv. Ale všechny vyžadují duševní dovednosti – jako vnímání, asociaci, predikci, plánování, řízení motoriky – které umožňují lidem a živočichům dosahovat jejich cílů. / Inteligence není jednorozměrná, ale je bohatě strukturovaným prostorem odlišných schopností pro zpracovávání informací. Proto používá AI mnoho různých postupů, určených k mnoha různým úkolům“ (Boden, 2016). Básní umělé inteligence má dvojí poslání, praktické, aplikované a teoretické, koncepční: „AI má dva hlavní cíle. Jeden je *technický*: používáním počítačů dosáhnout užitečných věcí (někdy za použití metod, které jsou velmi *odlišné* od postupů používaných myslí). Druhý je *vědecký*: používáním pojetí a modelů umělé inteligence napomoci k zodpovězení otázek týkajících se lidských bytostí a jiných živých záležitostí“ (Boden, 2016).

O inteligenci lze uvažovat jako o produktu a rámci předvídatelného jednání. Rozdíl mezi pouhým živočišným reagováním, byť založeným na nějakém druhu ukládání zkušeností dlouhodobé praxe a komplexnější schopnosti tvořit složité strategie, jejichž jednotlivé etapy nemusí mít vůbec bezprostřední, rozpoznatelnou souvislost s formulovaným cílem, je v tomto smyslu podstatný. Informace, které získáváme

z vypěstované citlivosti ke svému prostředí, se transformují do hypotéz a od aktuálního světa osvobozených alternativ. Je pravděpodobně příznačné, že posun k intelligentnímu chování provází především rozvoj distančních smyslových orgánů, které získávají převahu nad kontaktními. Důvodem může být jednak jejich vyšší strukturovanost, vyšší informační hustota, jednak vytvoření předpokladů pro uvolnění vazby podnětu a odezvy. Bouřlivý rozvoj zraku, jeho přímá vazba na mozek, pro který se zrakový nerv a oko stává vlastně „předsunutou“ součástí, provází osvobození organismu od přímé závislosti na aktuálních jevech, osvobození se od vlády přítomnosti vzhledu. Roste význam zprostředkovující informace, která začne být reprezentována znaky, využitelnými pro volné kombinování. „Vnitřní zrak“, směrování k představě, k obraznosti je presumpcí svobodného plánu, ovšem zároveň uvolněním pravdivosti. Odtud jistě také pramení spolehlání se silné AI na symbolickou reprezentaci.

Předvídatost člověka je významně většinou do intuice vidění. Z pouhého podhledu na strukturovaný celek spouští člověk řadu reakcí, které obsahují předběžná zhodnocení kontextových situací a jejich možného vývoje. Některé kognitivní teorie umění si všimají skutečnosti, že vnímání uměleckých děl dílem sítí a dílem zdokonaluje tuto schopnost vidění.

2 Modelování života

Sdílíme pragmatickou intuici, respektive neúplnou indukci, že každá inteligence je vázána na nějakou formu biologického života. A proto podobně jako vize umělé inteligence, vzniká i vize umělého života. Bodenová tuto souřadnost postihuje takto: „Umělý život (A-Life) modeluje biologické systémy. Jako AI obecně, má také jak technický, tak vědecký cíl. A-life je neodlučitelný od AI, protože veškerá inteligence, o které víme, se nalézá v živých organismech. A skutečně většina lidí věří, že mysl může vyvstat jenom z života [...]. Zatvrzelí technici se o tuto otázku nestarají. Ale obracejí se k biologii při rozvíjení praktických aplikací mnoha typů. Tyto aplikace zahrnují roboty, evoluční programování a samoorganizující se zařízení.“ Případně na jiném místě: „Všechny myslí, o kterých víme, se nalézají v živých organismech. Většina lidí – včetně kybernetiků [...] –

věří, že to tak *musí* být. Tedy domnívají se, že mysl nezbytně předpokládá život“ (Boden, 2016).

Předmětem modelování se tudíž stává i život. Jak ale život pojmet. Bodenová samozřejmě připomíná, že „[n]eexistuje všeobecně přijímaná definice života. Ale obvykle se uvádí devět rysů: samoorganizace, autonomie, emergence, růst, adaptace, citlivost, rozmnhožování, vývoj a metabolismus. Prvním osmi lze rozumět jako informačnímu zpracování, takže je v zásadě možné reálně je vytvářet prostředky umělé intelligence/umělého života. Například samo-organizace, která, jak se jí všeobecně rozumí, obsahuje všechny ostatní rysy, lze dosáhnout různými způsoby [...]. Ale metabolismus je odlišný. Může být *modelován* počítací, ale nemůže být počítací *reálně vytvářen, demonstrován*. Ani sebe-sestavující roboty, ani virtuální (obrazovkový) umělý život nemůže opravdu metabolizovat. Metabolismus je užití biochemických substancí a energetických výměn pro sestavování a udržování organismu.“ (Boden, 2016) Tak před námi vystavá potřeba adekvátně postihnout meze a prospěšnost počitačového modelování. Ze snahy dostatečně definovat život a využít k tomu počítací vyplývá tradiční antagonismus dvou jeho stránek – ideově-informační a materiálně-energetické.

V náhledu Ctirada Johna je „vhodným studijním modelem života v nejjednodušším a úplném provedení“ každá bakterie, protože „je schopná vykonávat všechny životní funkce“ (John, 1994). Bakterie je „systém s cílovým chováním“, přičemž lze rozlišit krátkodobý cíl – zachování existence buňky, a dlouhodobý – zachování kontinuity života, tedy bakterie neboli buňka, musí být zároveň systémem se sebeudržováním a systémem s autoreprodukcí.

Johnova redukce přináší návod na přijatelnou metodiku modelování života obsahující dve jeho základní stránky. *Systém se sebeudržováním* kontroluje výměnu látek a energií mezi systémem a okolím (prostředím), je otevřený, nicméně vede k ustálenému stavu. K udržení ustáleného stavu, k vyrovnávání odchylek slouží nejen patřičné senzory a zavedené strategie reakcí, ale i schopnost adaptace a učení. Učení je motivováno nejen zájmem organismu o zefektivnění vlastního chování, ale především dopadem proměňujících se vlastností prostředí. Základem je tu tedy metabolismus jako materiální stránka zahrnující růst, adaptaci a citlivost. Reakce umožňuje krátkodobá paměť zpracovávající aktuální informace. *Systém s autoreprodukcií* je především založen na stálé vnitřní paměti, která obsahuje informace pro zdvojení organismu. V tom smyslu jde vlastně o systém uzavřený. Genetická paměť, program pro reprodukční automatismus řídí samo-organizaci, informační stránku rozmnhožování a vývoj.

Pokud se dva objekty tvořené soubory dílčích prvků vzájemně podobají, je vhodné posoudit míru jejich podobnosti jako podobnost systémů, přičemž lze rozlišit podobnost chování a podobnost struktury. K jednomu z podobných systémů nalezneme systém

náhradní (zástupný), „[nazýváme ho] *modelem* a jeho využívání, nejčastěji však utváření, *modelování*.“ (John, 1994)

Je celkem přijatelné, že genetická informace, sestavovaná ze čtyř nukleotidů (adenin, thymin, guanin a cytozin), může být modelována na základě podobnosti systémů molekulární genetiky a lingvistiky, protože sdílejí podobnou architekturu, podobné principy konstrukce, týž hierarchický princip. Také lingvistika pracuje se subjednotkami – grafémy nebo fonémy, které nemají autonomii, nicméně na jejich základě vznikají slova, která už mají funkci autonomější. Nejvyšší autonomie nastává na úrovni syntaktické.

Kognitivní věda byla při svém zrodu inspirována počitačovou technologií. Díky definování abstraktních objektů, jejich vlastností a možných transformací plynoucích z obecných syntaktických pravidel se prostřednictvím počitačového programu mohou vytvářet struktury, které se počtem svých součástí, množstvím procesů a rychlostí opakování mohou přiblížit přirozeným sebeudržovacím a autoreprodukujícím projevům.

3 Samoorganizace, sebeuspořádání a růst komplexity (John von Neumann)

„Klíčovým rysem biologických organismů je jejich schopnost zkonstruovat sebe samotné. Samo-organizace je spontánní emergence řádu z počátku, který je uspořádán na nižším stupni. Je to matoucí, dokonce kvazi-paradoxní vlastnost. A není samozřejmé, že by se totéž mohlo dít s neživými věcmi. Celkem vypadá, samo-organizace je kreativním fenoménem“ (Boden, 2016).

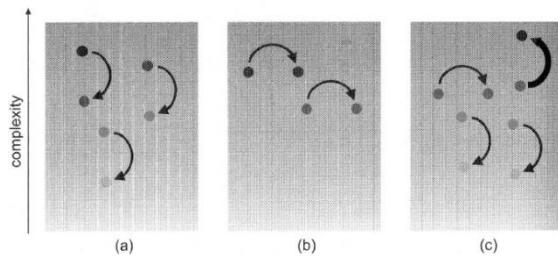
Pokud pro nás kreativita představuje užitečnou novost, je nově vzniklý organismus nebo stroj také účinný a uplatnitelný. Jeho existence se vrádí do stávajících kontextů a přivedí transformaci celku na novou úroveň. Kognitivní teorie umění totéž přisuzuje i hodnotným uměleckým dílům.

Pokud přisuzujeme v oboru umělé intelligence prorockou roli Turingovi, měl by být podobně za hlasatele umělého života pokládán von Neumann. Do svých úvah o počitačích záhy zařadil klíčový koncept dosud výhradně spojovaný s životem, a sice představu možnosti jejich sebereprodukce. K tomuto tématu se mnohokrát vrátil, a přestože byly tyto myšlenky za jeho života v tištěné podobě obtížně dostupné, pro svoji inspirativnost byly široce rozšířené, ať už zazněly na přednáškách, zachytily je jejich hojně kopírované záznamy nebo se sdílely ústním podáním.¹ V důsledku vážného onemocnění nechal svůj návrh sebereprodukčního strojů dopracovat Johna Kemenye (1955) a ten

¹ V průběhu 40. let; definitivnější verzi přednesl na Hixonském symposiu 20. září 1948 v Pasadeně (Kalifornie). Tato verze je zařazena do sebraných spisů Johna von Neumanna (1963), 5. díl, A. W. Taub (ed.). Kniha *Theory of Self-reproducing Automata* vyšla 9 let po von Neumannově úmrtí, v roce 1966 editována a dopracována A. W. Burksem.

později dále zjednodušil E. F. Codd (1968) (Kelemen, 2012).

V návaznosti na Turingovy stroje (automaty), které jsou pojímány jako posuvná zápisová páska se symboly, uvažoval von Neumann o strojích (automatech) takové konstrukce, jejichž produktem budou podobně jiné automaty. Principiálně tato myšlenka obsahuje problém stroje jako původce stoupající složitosti. Se stroji, které vytvářejí produkty nižší složitosti, se setkáváme běžně. Z informačního a termodynamického hlediska takový proces odpovídá degenerativním tendencím k entropii. Ještě se nám nezdá nemožné, aby stroj dokázal „kopírovat“ složitost vlastní. Při zahrnutí účinku náhodných mutací ovšem můžeme očekávat, že by některá „kopie“ mohla dosáhnout vyšší komplexnosti než její původce (Obr. 1). Takové úvaze byla samozřejmě blízká biologická praxe evoluce, tedy vznik složitějších organismů z jednodušších živých forem (Floreno a Mattiussi, 2008). Nalezení podobného postupu pro stroj plodící složitější stroj, než je sám, vyvolalo paralelu jednak součástky a orgánu, resp. buňky, jednak instrukce a genu.



Obr. 1: Abstraktní znázornění úlohy sebeprodukce pro růst komplexity strojů. Stroje jsou znázorněny teckami, šipky znázorňují výrobu. V případě (a) má vzniklý stroj nižší složitost, (b) je stejně složitý a v případě (c) lze na obrázku nalézt ojedinělý případ, kdy vznikl stroj složitější (Floreno a Mattiussi, 2008).

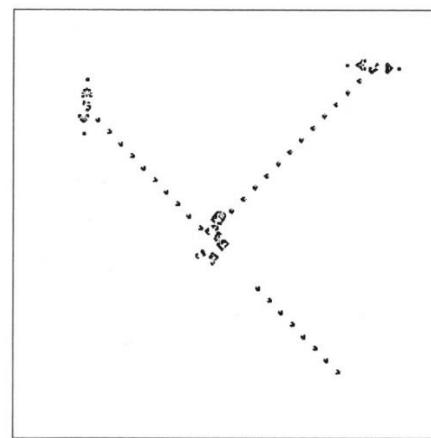
Od prvních představ, které ještě příliš konkrétně počítaly s motory, čidly a elektronickými prvky plovoucími po hladině „součástkové nádrže“, se von Neumann od tohoto kinematického návrhu přenesl na radu matematika Stana Ulama k buněčnému modelu, kde prostředím pro vznik samoreprodukujícího se stroje se stal dvourozměrný „celulární automat“. V návrhu zpracovaném v knize *Theory of Self-reproducing Automata* (1966) popisuje systém o 29 stavech s obecným pravidlem přechodů, uvádí konstrukci některých základních orgánů, konstrukci pásky a jejího řízení, konstrukci paměťové kontroly a univerzálního konstruujícího zařízení (Wünsch, 1968). Nulový stav, představující absenci hmoty v buňce, mohl přejít do několika běžných nebo zvláštních přechodových stavů, které mohou existovat jak v klidových nebo excitovaných podmínkách, případně se transformovat do dalších pomocných stavů. Přechodové funkce byly definovány tak, aby umožňovaly vytvoření a zničení „hmoty“ v celulárním prostoru místo jejího přenášení, jak tomu bylo u kinematického modelu. Von Neumann tak definoval

základní představu o smyslu celulárního, buněčného modelu (CA).

4 Modelování života celulárními automaty

4.1 Conwayova redukce CA

Po zmíněném Coddově modelu se objevil ještě více zjednodušující celulární automat modelu života Johna H. Conwaye. Dlužno přiznat, že byl uveden v kontextu sofistikované zábavy. Oč byl princip jeho „Hry života“ elementárnější, o to se stal pro souhru jednoduchých principů a působivou demonstraci výsledného nepřehledného chování přitažlivější. Na ploše je pravidelný pravoúhlý rastr buněk, stav buňky může být jen dvojí, odpovídající „životu-hmotě“ a „smrti-prázdnou“. Logicke pozadí transformací se smršťuje na vyhodnocování přítomnosti množství „hmoty“ v okolí buňky. Princip smíšení pouhého pohybu buňky a jejího vzniku či zániku působí sice vizuálně efektně, ale neumožňuje sledovat, případně rozvíjet genetické strukturování a jen velmi omezeně energetické, hmotné okolnosti. Model je to tak živý, jako těsto v mechanickém hnětači.

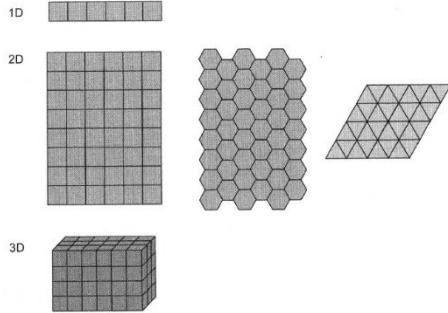


Obr. 2: Kluzákové dělo. Jedna ze situací „Hry života“, kdy souhra periodického vznikání evokuje iluzi biologického zrodu organismu a jeho pohybu (Coveney a Highfield, 2003).

4.2 Obecný tvar buňky CA

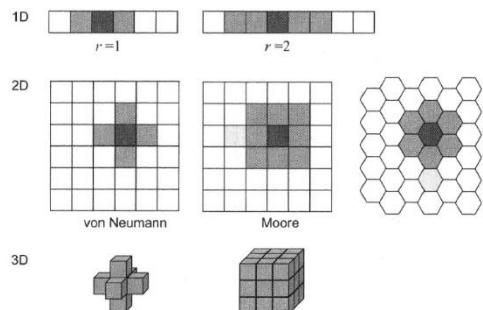
Pro další promýšlení vizuálních možností buněčného automatu a zároveň možnosti jak organizovat logické předpoklady jeho chování není samozřejmě bez významu ani samotný tvar buněk, ani jejich rozměrová expanze. Pravoúhlá mřížka je sice pohodlnější zvládnutelná karteziánským adresováním, ale rastr rovnoramenných trojúhelníků nebo pravidelných šestiúhelníků také bezezbytku pokrývá plochu, ale navíc i umožňuje snadněji zahrnout do úvah souvislost

konkrétních vzdáleností středů buněk a vrstev jejich sousedství.



Obr. 3: Vyhodnocování vztahů jednotlivých buněk může být redukováno na jeden rozměr, nebo rozšířeno na tři rozměry, přičemž samo definování tvaru buňky ovlivní významně možnost logických dispozic transformací (Floreno a Mattiussi, 2008).

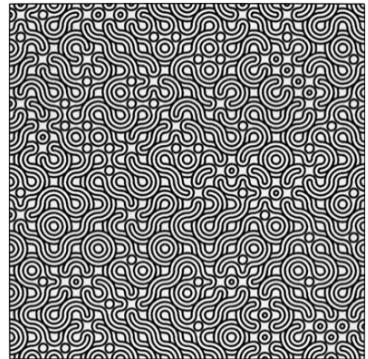
Prostorové modelování buněčného automatu je blíže naši přirozené zkušenosti, ale vzhledem k praxi vidění vzniká problém s nedostupností většiny informace, případně praktickým provedením trojrozměrné simulace. Plošná abstrakce naopak může těžit ze zkušenosti značné části výtvarného umění.



Obr. 4: Působení sousedství na konkrétní buňku může vyplývat z charakteru dělení prostředí, může však zahrnovat i parametr vzdálenosti, který lze interpretovat buď jako prostorovou nebo prostorově-časovou okolnost (Floreno a Mattiussi, 2008).

4.3 Plošný CA model v umění (Paul Brown)

Dlouhodobě se využitím CA modelování života inspiroval Paul Brown, britský umělec neokonstruktivního zaměření. Od rigidního čtvercového rastru dospěl k evokaci fluidní, přelévací podstaty života, protože původní čtvercový rastr rafinovaným způsobem překryl způsoby vizuálního sdružování sousedních tvarů. Také rozmnovení typů základních prvků může být užitečné pro postihování složitějších etap vývoje než zánik a zrod.



Obr. 5: Paul Brown: Bez názvu, 1975. Počítačem podporovaná kresba.

4.4 Prostorový CA model v umění (Yoichiro Kawaguchi)

Na počátku 80. let se s přispěním programátora Eiichi Izuhary podařilo Yoichiro Kawaguchimu uspět s vytvořením růstového buněčného algoritmu na prostorovém základě². Cesta k programu GROWTH vedla od lineárně kreslicího počítačového systému z roku 1977 přes vyvíjení síťových kreseb ulit a rostlin, později s užitím fraktálového opakování. Nové stadium simulace života si vzalo za základ třírozměrný buněčný prostor, přičemž jeho variace lze chápat buď jako časové vrstvy, nebo jako sousedství buněk. Kawaguchi zahrnul do programu oba způsoby, přičemž je opět základním pravidlem hustota sousedství živých buněk v okolí, která ovšem mohla průběžně kolísat (Kawaguchi, 1997 a Kawaguchi, 1982).



Obr. 6: Yoichiro Kawaguchi: Buňka metropole umělého života, 1993.

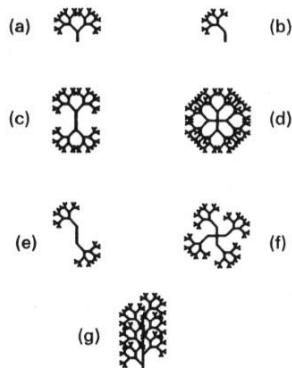
² Svoji simulaci života Kawaguchi úspěšně prezentoval na umělecké konferenci a výstavě SGGRAPH '93.

5 Genové modelování

5.1 Dawkinsovy biomorfy

Dawkinsovo modelování evolučního procesu z roku 1986 abstrahovalo jak od energetického, metabolického pozadí životních procesů, tak od ontogenetického cyklu každého organismu. Základním principem je segmentovaná struktura celé řady běžných živočichů. Dawkinsovy zástupné organismy, biomorfy, tvoří spojení úseček sestavovaných stromovým větvením při rekursivním naprogramování (Dawkins, 2002). Počet propojených součástí neboli řetěz opakování odpovídá počtu rekurzí, které musí být dopředu stanoveny. To určuje první z devíti genů užitých pro toto modelování. Každý gen má vždy nějakou konkrétní hodnotu ve stanoveném rozsahu. Další pozice genové informace ovlivňují úhel přírůstku, jeho velikost, vzdálenost takto vzniklých segmentů, průběžný tvarový nárůst segmentů atd.

Dawkinsovy biomorfy velmi názorně modelují vztah informace zakódované do souboru hodnot, který se programem snadno převede na viditelný výsledek morfogeneze.



Obr. 7: Uplatnění genů různých druhů symetrie na segmentový organismus v Dawkinsově evolučním programu. (Dawkins, 2002).

5.2 „A-Volve“ (Christa Sommerer a Laurent Mignonneau)

Evoluční prostředí umělého života nazvané „A-Volve“ vytvořila na základě inspirace Dawkinsovými biomorfami umělecká dvojice Christa Sommererová a Laurent Mignonneau v letech 1994-1995. Jedná se o interaktivní počítačovou instalaci pracující v reálném čase. Jejím základem je dotyková obrazovka, která ve vodorovné poloze zobrazuje virtuální organismy. Tito virtuální tvorové vznikají na základě genetického kódu, který se přenáší mezi generacemi tvorů a který určuje jejich tvar a velikost. Každý tvor může zaujmít dvojí chování. Bud' se stává aktivním lovčem, nebo naopak unikající kořistí. Tato vlastnost je umožněna zahrnutím informace o energetických výdajích a potřebách vir-

tuálních živočichů (Sommerer a Mignonneau, 1997). Vytvořený program modeluje schopnost tvora získávat informace v určitém výseku okolí, v jeho „zorném poli“, dále modeluje možnost získat potravu z jiného živočicha, ukazuje individuální růst organismu a také jeho „sexuální“ chování, které produkuje nové jedince nesoucí průnik genetických informací rodičů.

6 Parametry návrhu organismu uměleckého díla

Uvažovaná paralela uměleckého díla a organismu (resp. ekosystému) se samozřejmě nemůže opírat o umění duchampovsky subversivní nebo o umění konceptuálně odvozené od ideologických, psychologických a psychologických kalkulací, ale musí se vrátit k tradici umění modernistického. Dlouhodobý vývoj prověrující možnosti strukturace obrazového prostoru v kompozici nebo konkretistické strukture vyvrcholil právě ve výtvarném umění první poloviny 20. století. Nahlížel tradiční stavební prvky vizuálního umění jako praxi spojování jednodušších prvků do stále propojenějších, hierarchizovanějších a složitějších celků. Cílem je poskytovat dostatečné vizuální klíče přehlednosti a zároveň dodávat širokou škálu možných vizuálních zážitků.

Ve smyslu Johnova návrhu (viz kap. 2) lze uvažovat o paralele následujících vlastností života v rámci ekosystému a tvorby výtvarného díla v modernistickém smyslu.

Organismus a ekosystém

- Individuální růst, podporovaný metabolismem a reagování na prostředí (ontogeneze)
- Vývoj organismů (fylogeneze) a jejich sociální interakce; růst komplexity
- Fitness jako zajištěné dynamické trvání, setrvalý rovnovážný stav

Umělecké dílo

- Propracovávání vstupních elementů, jejich sestavování, sdružování
- Zahrnutí vzájemného opakování působení do strukturovaného celku; růst komplexity
- Fitness jako dynamický, ale setrvalý a rovnovážný strukturovaný celek (jednotlivé dílo, cyklus, osobní autorský styl a umělecký směr)

Modelování organismu v ekosystému, resp. uměleckého díla bude muset vycházet z energetické bilance celku, která by měla mít stále stejnou hodnotu. Jednotlivá stadia dílčích organismů by měla vznikat opakováně a proměňovat se vlivem vzájemného působení. Kromě procesů ontogeneze by vzniklé součásti měly vytvářet vnitřní paměť svého rozvoje, která by podporovala opětovný výskyt jednotlivých řešení, jejich sebereprodukci. Kromě toho musí být tato vnitřní informace o postupu vlastní výstavby, o způsobu strukturování, do určité míry variabilní. Přípustnost nepřesné sebereprodukce uvolní varianty, které průběž-

ně omezí interakce s prostředím, tedy s postupně tvořeným celkem. Reaktivnost organismu na prostředí musí zahrnovat informace o větším než bezprostředním sousedství, přičemž ideálně bude směrovat k rozpoznaní nejen aktuálního stavu buněčných celků, ale i jejich intence.

Jinak shrnuto, při řešení počítačové simulace života by mělo být komplementárně doplňované zvětšování počtu diferencujících se prvků možností jejich slučování do jednotek vyšší vrstvy, při stoupajícím způsobu jejich interakce. Vizuální složitost lze běžně postihnout v kategoriích obecných typových tříd, „znakových komplexů“ neboli „superznaků“, které mají sice arbitrární důvod ustavení, ale zároveň vůči sobě statisticky měřitelný rozdíl rozpoznatelných vlastností obsahu. Růst složitosti organismu by měl modelovat i kapacitu vhodnou pro symbolickou manipulaci, zakládající uměleckou účinnost. Zjednodušeně řečeno, kvantitativní násobení prvků struktury musí provázet rozvíjení vnitřní hierarchie, tedy kvalitativní růst. A právě v těchto kategoriích se tradičně uvažuje i o struktuře uměleckého díla.

Poděkování

Tento příspěvek vznikl za podpory programu PROGRES Univerzity Karlovy v rámci dílčího programu Q21 Text a obraz ve fenomenologii a sémiotice pod číslem projektu FHS/Q21_206021/28.

Literatura

- Bertalanffy, L. von (1972). *Člověk-robot a myšlení, Psychologie v moderním světě*. Svoboda.
- Boden, M. A. (2016). *AI, Its nature and future*. Oxford University Press.
- Coveney, P. a Highfield, R. (2003). *Mezi chaosem a rádem, Hranice komplexity: hledání rádu v chaotickém světě*. Mladá fronta.
- Dawkins, R. (2002). *Slepý hodinář, Zázrak života očima evoluční biologie*. Paseka.
- Floreno, D. a Mattiussi, C. (2008). *Bio-Inspired Artificial Intelligence, Theories, Methods, and Technologies*. Massachusetts Institute of Technology.
- John, C. (1994). Modely v biomedicínském myšlení. Ve sborníku *Model a analogie ve vědě, umění a filozofii*. Filosofia, str. 125-138
- Kamarýt, J. (1994). Analogie a homologie jako nástroj poznání a omylu. Ve sborníku *Model a analogie ve vědě, umění a filozofii*. Filosofia, str. 139-158

Kawaguchi, Y. (1982). A Morphological Study of the Form of Nature. *Computer Graphics* 3(16): 223-232

Kawaguchi, Y. (1997). The Art of the GROWTH Algorithm with Cells. Ve sborníku *Artificial Life V*, str. 159-166

Kelemen, J. (2012). *Kyberkreativita, Stroj, tvorivost, člověk*. Slezská univerzita v Opavě.

Sommerer, Ch. a Mignonneau, L. (1997). „A-Volve“ an evolutionary artificial life environment. Ve sborníku *Artificial Life V*, str. 167-175

Wünsch, Z. (1968). John von Neumann, Theory of Self-reproducing Automata. *Kybernetika* 5(4): 495-496.

Role třídění z hlediska stability v evoluci

Jan Toman, Jaroslav Flegr

Katedra filosofie a dějin přírodních věd, Přírodovědecká fakulta, Univerzita Karlova
Viničná 7, 128 00 Praha 2, Česká republika
tomanj@natur.cuni.cz
flegr@cesnet.cz

Abstrakt

Přirozený výběr bývá považován za hlavní proces pohánějící biologickou evoluci. Ve skutečnosti se však jedná o manifestaci obecnějšího „perzistentního principu“, důsledky jehož činnosti v historických systémech jsme navrhli nazvat „třídění z hlediska stability“ (stability-based sorting, SBS). Na rozdíl od oportunistického výběru se mohou prostřednictvím SBS akumulovat i momentálně nevhodné vlastnosti prospěšné pro dlouhodobé přežívání tříděných entit. Není sice tak rychlý a nedokáže vytvářet komplexní adaptace, v evoluci má však vždy poslední slovo. SBS implikuje například nárůst komplexity v evoluci, dokáže ale vysvětlit i řadu dalších záhadných fenoménů biologické, kulturní a umělé evoluce. Může se jednat o jediný zcela univerzální evoluční proces a jeho docenění může přinést nový výhled do evoluce komplexních systémů.

1 Úvod

Přirozený výběr, zřejmě nejvýznamnější objev Charlese Darwina (Darwin, 1859), nabídl přesvědčivé vysvětlení vzniku a hromadění adaptivních, nezřídka sofistikovaných a komplexních, znaků organismů. K jeho průběhu stačí, aby entity v systému vznikaly ve vzájemné závislosti, například kopírováním či množením, směrem k variabilnímu potomstvu vykazovaly určitou míru dědičnosti a měly nadprodukci potomků. Vysvětlení a řešení založená na principu přirozeného výběru byla od poloviny 19. století aplikována snad ve všech oborech přírodních, technických i humanitních věd. Evoluční biologové v průběhu času ukázali, že selekce má řadu komponent a řadu forem, a že mimo ni pohání a výrazně ovlivňuje vývoj organismů i mnoho jiných procesů, jako jsou genetický drift, genetický draft, evoluční tahy, genový tok či druhový výběr (viz např. Mayr, 2003; Flegr, 2005). Rovněž se ukázalo, že řada adaptivních vlastností primárně nevznikla jakožto biologické adaptace, ale jako exaptace, tj. přizpůsobení, která původně sloužila k úplně jiné funkci, nebo spandrel, tj. strukturní či funkční prvky organismů bez konkrétní funkce vynucené čistě vnitřní organizací živých bytostí,

a svou pozdější funkci získala až dodatečně (viz např. Gould, 2002). V poslední době se navíc do hledáčku evolučních biologů vrátila provázaná problematika komplexní podstaty genetické dědičnosti, různých forem negenetické dědičnosti a mnohoúrovňových meta-adaptací ovlivňujících evolvabilitu evolučních linií, tj. schopnost generovat adaptivní varianty, a kanalizujících jejich evoluční změny – zejména individuálního vývoje mnohobuněčných živočichů (viz např. Laland a kol., 2015).

Relativně pozdní objev přirozeného výběru bývá považován za jednu z největších záhad historie vědy. Tento z dnešního pohledu naprostě evidentní proces byl popsán až v druhé polovině 19. století, tedy daleko později než většina srovnatelně složitých nebo i daleko složitějších jevů v jiných vědeckých oborech (Komárek, 2004). Jedním z navržených vysvětlení této záhady je koncept tzv. sociomorfního modelování (Komárek, 2008). Podle něj mohl být Darwinův model selekce vysvětlující evoluci organismů jako důsledek soupeření jedinců o co nejvyšší zdatnost vytvořen a všeobecně nahládnut teprve v Anglii 19. století, ve které obdobné soupeření mezi jednotlivými ekonomickými subjekty vedlo k nesmírně rychlému a nápadnému vývoji v průmyslové výrobě a společnosti. S ohledem na to není náhodou, že prakticky identické modely evoluce prezentovali Matthew (1831), Darwin (1859; Darwin a Wallace, 1858) a Wallace (Darwin a Wallace, 1858) v rozmezí pouhých několika let.

Koncepty uvažující o historických proměnách organismů, a tedy zahrnující určitou verzi biologické evoluce, se přitom objevovaly už od předsokratického období starověkého Řecka (jak shrnuli Hladký, Kočandrle a Kratochvíl, 2012). Empedokles, například, postuloval ve své zoogonii vznik organismů prostřednictvím náhodného kombinování různých údů (orgánů). Někteří takto vzniklí tvorové byly neživotaschopné zrůdy. Ani většina ostatních nebyla příliš úspěšná, protože k sobě jejich části příliš nepasovaly. Takové organismy postupně vymřely. Některé z organismů – takové formy, se kterými se setkáváme dnes – se nicméně ukázaly jako dobré organizované, úspěšné a za přirozených podmínek perzistentní, takže v přírodě převážily. Podobné třídění vznikajících organismů na základě jejich

životaschopnosti v daných podmírkách uvažoval i Lucretius.

Bыло бы jednoduché tyto zvláštní aspekty starověkých modelů evoluce odvrhnout jako pouhé nepochopení přirozeného výběru. Tím bychom ale patrně jen utvrdili nadvládu sociomorfního modelování nad naším uvažováním. Starověké autory, jejichž společnost se za jejich životů příliš neměnila a kteří vnímali čas spíše jako cyklický fenomén, totiž s největší pravděpodobností jejich zkušenosť přivedla k objevu a docenění univerzálního a dnes trestuhodně opomíjeného evolučního procesu, který ovlivňuje všechny materiální, imateriální, živé i nežitě entity v systémech podléhajících evoluci a jehož je přirozený výběr jen jedním specifickým případem: třídění z hlediska stability (stability-based sorting, dále SBS) (Flegr a Toman, 2017).

2 Druhy stability v evoluci

2.1 Přežití stabilnějšího

Ačkoli se tento fakt běžně nekomentuje, opakováně bylo navrženo, že je přirozený výběr manifestací obecnější zákonitosti, která ovlivňuje všechny materiální i nemateriální entity ve vesmíru a která, na rozdíl od selekce, nevyžaduje replikaci v závislosti na ostatních entitách, dědičnost vlastnosti ani nadprodukci variabilního potomstva. Nejběžněji bývá tato zákonitost nazývána „přežití stabilnějšího“ (survival of the stable) podle postřehu Richarda Dawkinse v knize Sobecký gen (Dawkins, 2003, str. 22-23): „Darwinovo „přežití zdatnějšího“ je v podstatě speciálním případem obecnějšího pravidla přežití stabilnějšího. (...) První formou přírodního výběru byl výběr stabilních struktur a zamítnutí nestabilních. Není v tom žádná záhada. Stalo se to zákonitě.“

Postřeh, že nestabilní entity přicházejí a odcházejí, zatímco stabilní nebo neustále vznikající se v systému hromadí a převažují, respektive že nejnestabilnější (nejperzistentnější) entita vydrží nejdéle, je nepochyběně axiomatický (Pross, 2012; Shcherbakov, 2012; Pascal a Pross, 2015). Nejedná se tudíž o žádnou hypotézu, ale samopotvrzující se výrok. To v situaci, kdy nás zajímá konkrétní průběh evoluce systému, není takový problém¹. Problematičtější jsou evoluční

důsledky tohoto procesu, které se mohou na první pohled zdát poměrně triviální. V reálném světě ovšem dochází ke komplexním interakcím mezi koexistujícími entitami, které oplývají mnoha vlastnostmi, jejichž vliv na perzistence entit se liší a může být kontextově závislý. Vývoj takovýchto systémů, mezi které spadá i biologická evoluce – tedy to, co nás zajímá a co zkoumáme – tak rozhodně triviální být nemusí (Bardeen, 2009; Pross, 2012; Wagner a Pross, 2011; Pascal a Pross, 2015; de Vladar a kol., 2017).

Obecné postřehy analogické Dawkinsovu „přežití stabilnějšího“ učinili i další autoři (viz např. Lotka, 1922; Simon, 1962; Van Valen, 1989; Pross, 2012; Maynard Smith a Szathmáry, 2010; Wagner a Pross, 2011; Pascal a Pross, 2015). Stejně tak byl fenomén ve svých konkrétních podobách popsán pod různými jmény v řadě oborů, které se zabývají různými formami biologické i nebiologické evoluce (viz např. Slobodkin a Rapoport, 1974; Rappaport, 1999; Pross, 2012; Bouchard, 2008, 2011; Flegr, 2010, 2013; Wagner a Pross, 2011; Shcherbakov, 2012; Bourrat, 2014; Doolittle, 2014; Pascal a Pross, 2015; Bardeen a Cerpa, 2015). V oboru evoluční a teoretické biologie je mu však tradičně příkládána spíše malá důležitost. Například Okasha (2006, str. 214) či Godfrey-Smith (2009, str. 40 a 104) nepovažují fenomén za evolučně zajímavý, protože podle nich jeho působením, na rozdíl od působení paradigmatického přirozeného výběru, nemohou vznikat zajímavé adaptace. Opačný názor je více méně okrajový (viz např. Bouchard, 2011; Doolittle, 2014; Bourrat, 2014).

Autoři tohoto příspěvku jsou přesvědčení, že tento univerzální evoluční proces „přežití nejnestabilnějšího“, který je v současnosti nedoceňován a často není vůbec brán v úvahu při snaze vysvětlovat vznik konkrétních znaků a vlastností živých organismů, může být ve skutečnosti odpovědný za celou řadu adaptací – zejména takových, jejichž vysvětlení působením přirozeného výběru je považováno za problematické až nemožné. Kvůli jeho slabému koncepтуálnímu ukotvení jsme potom navrhli tento proces, tj. obecnou tendenci systémů k hromadění stabilnějších, perzistentnějších a ke změně rezistentnějších entit a jejich vlastností, nazvat třídění z hlediska stability (Flegr a Toman, 2017).

2.2 Statická a dynamická stabilita

V drtivé většině systémů probíhá SBS jakožto kompetice mezi entitami o co největší statickou stabilitu, tj. o co nejmenší pravděpodobnost vlastního zániku nebo vlastní transformace v něco jiného. V systémech, kde nové entity vznikají z entit původních a dědí přitom jejich vlastnosti, se soutěží nikoli o statickou stabilitu, ale o stabilitu dynamickou (Pross,

¹ Situace je to ostatně analogická diskuzím, zda je biologická zdatnost (nebo ve fyzice hmotnost) definována kruhem. Ať už ano, nebo ne, není to vlastně podstatné. Stejně jako zdatnost, i stabilita (perzistence) potenciálně závisí na mnoha různých vlastnostech a kontextech, kterými entita prochází. Jedná se tak vlastně pouze o technický, zastřešující termín pro vlastnosti, které za daných podmínek zvýhodňují určité entity v průběhu třídění z hlediska stability na úkor jiných. Jednou z nich může být nízká hladina entalpie, další schopnost replikace, a ještě jinou třeba schopnost

určitých memů přetrvat a šířit se v kulturní evoluci. Hlubší diskuse o tomto tématu jde nad rámec příspěvku, můžeme ale odkázat např. na Flegra (2005).

2012; Wagner a Pross, 2011; Pascal a Pross, 2015). V těchto případech se vlastně emancipuje informace od hmoty (Shcherbakov, 2012) a kompetice stabilně existujících entit o co nejdéle statickou výdrž se mění na kompetici o to, které za dobu své existence dokáží vyprodukovať co největší počet svých kopií (respektive kopií informace určující vytváření vlastních kopií), nebo přesněji, která má nejvyšší rozdíl rychlosti vznikání a rychlosti zanikání svých kopií. Tento rozdíl pochopitelně závisí nejen na době jejich existence (u organismů na délce jejich reprodukčně aktivního života), ale i na rychlosti jejich reprodukce.

Všechny formy selekce, včetně darwinovského přirozeného výběru (Darwin, 1859), hamiltonovsko-dawkinskovského mezialeického výběru (Dawkins, 2003) či druhového výběru (Gould, 2002), jsou tak vlastně speciálními případy třídění z hlediska stability ve strikném smyslu a obvyklém pojednání, tj. třídění z hlediska statické stability. Vzhledem k tomu, že oba procesy mají výrazně odlišné evoluční konsekvence, které v tomto příspěvku chceme demonstrovat, ovšem má smysl mezi oběma procesy striktně rozlišovat. Třídění z hlediska statické stability, tj. třídění z hlediska stability v úzkém slova smyslu a obvyklém pojednání, proto budeme nadále označovat jednoduše jako třídění z hlediska stability. Třídění z hlediska dynamické stability nadále budeme nazývat jeho obvyklým názvem selekce.

Podobné vymezení je v souladu s tím, které zvolili ve svých studiích např. Slobodkin a Rapoport (1974), Rappaport (1999), Bourrat (2014), Bouchard (2008, 2011) či Bardeen (2009). Do největší hloubky se problému stability v přírodě a jejímu rozdělení na různé typy patrně věnoval Addy Pross se svými kolegy (viz např. Pross, 2012; Wagner a Pross, 2011; Pascal a Pross, 2015). K problému ovšem přistoupili z jiného úhlu, konkrétně s cílem objasnit chemicko-fyzikální podstatu živých organismů a vzniku života, takže se jejich pojednání různých druhů stability od našeho v důležitých detailech odlišuje. Zatímco nás „evolučně-biologický“ či „systémový“ přístup nebene v potaz fyzikální podstaty tříděných entit ani fyzikální bázi různých druhů třídění, ale pouze jejich perzistence v čase, Pross se svými kolegy se právě na tento aspekt stability zaměřují. Rozlišují statickou termodynamickou stabilitu, která se uplatňuje u neživých materiálních entit, a dynamickou kinetickou stabilitu, která je založená na replikativní chemii a je charakteristická pro živé entity, subjekty biologické evoluce. Oba druhy stability, stejně jako potenciálně další druhy stability založené na jiných základech, je podle nich možné shrnout pod hlavičku čistě logického perzistentního principu – obecné tendenze systémů měnit se od méně stabilních (méně perzistentních) forem k stabilnějším (více perzistentním) (Pascal a Pross, 2015).

3 Důsledky třídění z hlediska stability pro průběh evoluce

SBS i selekce mohou probíhat v uzavřených i otevřených systémech. Zatímco je ale k fungování jakékoli formy přirozeného výběru třeba, aby entity v systému vznikaly ve vzájemné závislosti, například kopírováním či množením, vykazovaly určitou míru dědičnosti a měly nadprodukci vzájemně variabilních potomků (Gould, 2002; Flegr, 2005, 2015; Okasha, 2006; Godfrey-Smith, 2009), SBS nic takového nevyžaduje. K třídění na základě stability dochází ve všech systémech podléhajících evoluci, tj. systémech s pamětí či historií v širokém slova smyslu. Podléhají mu všechny materiální i nemateriální entity bez ohledu na jejich původ, včetně entit, které vznikají nezávisle na sobě (např. sněhové vločky, kosmické objekty, memy či vzájemně izolované civilizace).

Velmi důležitou otázkou, přinejmenším z teoretického hlediska, jsou evoluční důsledky třídění z hlediska stability. Bylo by naivní předpokládat, že se z jeho vlivu živé entity podléhající biologické evoluci založené na spektakulárně se projevujícím přirozeném výběru zcela vymanily. SBS se projevuje ve všech systémech podléhajících evoluci, neustále a na všech úrovních. Shcherbakov (2012) na tomto základě odvozuje, že nevyhnutelným důsledkem každé evoluce, tedy i biologické, je stáze – stav, kdy se už entity nebudou dále vyvijet, měnit. Jakékoli evoluční změny organismů, jejich diverzifikace a komplexifikace, jsou podle tohoto přístupu vlastně jen vedlejšími produkty evoluce hledající nejstabilnější stav, důsledkem nemožnosti zcela eliminovat vznik mutací nebo projevem dočasného převládnutí oportunistického výběru preferujícího entity s vyšší dynamickou, ale nižší statickou stabilitou. Podobně triviální průběh může mít i evoluce, ve které absolutně převládne tlak na dynamickou stabilitu v její nejjednodušší podobě, tj. rychlosť množení. To se týká zejména různých pokusů o počítacové či laboratorní simulování otevřené (open-ended) evoluce (viz např. Spiegelman a kol., 1965; Ray, 1993, 1997; Thearling a Ray, 1994, 1996; Yaeger, 1994; Oehlenschläger a Eigen, 1997; Ray a Hart, 1998; Sayama, 1999; Adami a kol., 2000; Channon, 2001; Suzuki a kol., 2003), které patrně kvůli omezenému množství interakcí mezi koevolvujícími organismy a nedostatečně heterogenním prostředím zpravidla rychle vedou k vyselektování co nejjednodušších a co nejrychleji se množících entit. Wagner a Pross (2011) a Pross (2012) zastávají opačný názor než Shcherbakov (2012). Podle těchto autorů představuje prvek statické stability, termodynamická stabilita, pouze rámec, ve kterém se musí systémy replikujících se entit, například organismů v průběhu biologické evoluce, pohybovat. Na své nižší úrovni ale zároveň vykazují tendenci k rozrůznění a komplexifikaci.

Autoři tohoto příspěvku jsou na rozdíl od výše uvedených autorů přesvědčeni, že se SBS v systémech replikujících se entit přímo uplatňuje, ale jeho role je

spíše subtilní, přičemž selekce je spíše nástrojem třídění z hlediska stability než jeho pouhým vedlejším produktem. V jednoduchém, stabilním a homogenním prostředí by entity soupeřily pouze o nejvyšší počet co nejpřesnějších kopií, tj. rychlosť a přesnost reprodukce, které se zpravidla projevují jejich redukcí až na nezbytné minimum. V reálném světě ovšem na koexistující entity působí daleko heterogennější podmínky prostředí včetně ostatních koevolvujících entit, které dynamicky reagují na evoluční protitahy ostatních. Výsledkem je neustálé napětí mezi tlakem na konzervaci informace (tj. zvýšení rychlosti množení a přesnosti replikace, nebo zvýšení statické perzistence) a její evoluci (tj. adaptaci na nové podmínky). Za takových podmínek nemusí být nejúspěšnější entity, které se nejrychleji a nejpřesněji množí. Nejvyšší stabilita sice zůstává ultimátním atraktorem, ale organismy se ji nedobírají triviálnimi cestami (zvýšení statické perzistence, nebo rychlosti a přesnosti množení), ale sofistikovanými adaptacemi reagujícími na momentální kontext živého a neživého prostředí. Evolvabilita, v tomto smyslu schopnost podléhat adaptivní evoluci, tak může být důležitou metaadaptací umožňující zvyšování stability entit během selekce. Samotné interagující entity, respektive jejich evoluční linie uchovávající informaci, jak jedince replikovat, se tak sice v průběhu evoluce postupně mění, ale stále zachovávají historickou individualitu – stále jde o třídění z hlediska stability. Výsledný stav systému umožňuje odhadnout teorie her, respektive teorie evolučně stabilních strategií (Maynard Smith a Price, 1973), přičemž platí, že při dostatečné komplexitě prostředí a míře interakcí mezi entitami nemusí systém bezprostředně spět ke stázi ani převaze entit s nejvyšší rychlostí a přesností reprodukce, ale naopak rozrůznění a komplexifikaci entit (de Vladar a kol. 2017).

Na pozadí otevřené evoluce ovšem i tak dochází vlivem třídění z hlediska stability k hromadění entit, které odolávají působení přirozeného výběru a mají sníženou či zcela redukovanou evolvabilitu – takové entity, případně vlastnosti, se mohou jen hromadit, takže v průběhu evoluce přibývají. Pokud se ukáže ireverzibilní „zamrzlost“ dané vlastnosti výhodná, či nevratně „zamrzlá“ entita perzistentní za širokého rozmezí podmínek, může v daném systému zcela převážit. Evoluci v největším časovém měřítku tak lze číst jako neustálý soubor mezi stabilitou a konzervací na jedné straně a adaptací na druhé. Jak uvidíme dále, nelze přitom vyloučit, že právě tento charakter evoluce může být zodpovědný za mimořádnou bohatost a komplexitu pozemských organismů.

4 Vznik fenoménů působením třídění z hlediska stability

4.1 Rozdíly mezi působením selekce a SBS v evoluci

V systémech podléhajících přirozenému výběru zanechávají nejúspěšnější jedinci, tj. jedinci, jejichž vlastnosti nejlépe odpovídají momentálním požadavkům prostředí, před svou smrtí největší počet potomků, kteří se dožijí dospělosti a vyprodukují další generaci. V systémech podléhajících SBS jsou nejúspěšnější entity ty nejstabilnější, tj. takové, které vydrží nejdelší dobu před svým zánikem nebo změnou v něco jiného. Stabilita vznikajících tříděných entit je přitom, pokud neuvažujeme další faktory, zcela nezávislá na stabilitě entit již v systému přítomných. Selekce je tak daleko efektivnější, neboť díky dědičnosti vlastností, variabilitě potomků a faktu, že počet výhodných vlastností přímo ovlivňuje úspěch jedince, umožňuje hromadění a vylepšování adaptivních vlastností. V systému přibývají lépe přizpůsobené entity s větší biologickou zdatností, a naopak ubývají ty méně přizpůsobené.

Třebaže je tedy SBS rozšířenější, prostřednictvím selekce vzniklo v biologické evoluci nesouměřitelně větší množství nezřídka i velmi komplexních adaptací. Přesto je velmi pravděpodobné, že určitá třída adaptací, se kterými se setkáváme u dnešních organismů, vznikla působením SBS. Jedná se o jednoduché znaky, které vznikly jednou či dvěma změnami, vesměs s následnou rolí adaptace zcela nesouvisejícími přizpůsobeními. Důležitým zdrojem adaptací ovlivňujících stabilitu tříděných entit jsou zejména exaptace a spandrelы. SBS v tomto ohledu funguje jako síto, které vybírá vlastnosti zvyšující dlouhodobou stabilitu entit tvořících systém i systému samotného (Doolittle, 2014).

Typickým příkladem takové vlastnosti mohou být například různé na hustotě populace závislé metody redukce množení – teritorialita, sociální hierarchie nebo citlivost některých organismů k vyrušování během páření či péče o mláďata (Wynne-Edwards, 1986; Dawkins, 2003). Ať už je proximální role těchto adaptací jakákoli – třeba i vůbec žádná, nebo svému nositeli dokonce škodí – v širším pohledu, a tedy z hlediska SBS, se jedná o jakési „pojistky“ proti přemnožení populace. Populace a druhy, které mají tyto „pojistky“, netrpí často fatálními následky přemnožení, mají nižší riziko vymření a v delším časovém měřítku převáží nad populacemi a druhy, které vyštěpují sobecké mutanty schopné množení za jakýchkoli podmínek. Po nějaké době se budeme setkávat jen s potomky těchto populací a druhů, které měly výhodu v druhovém výběru. Akumulace neodbouratelných „pojistek“ působením SBS je v tomto ohledu naprostě zásadní. Skupinový (resp. druhový) výběr je totiž slabší než individuální a čistě jeho působením se patrně nemohou fixovat vlastnosti nevhodné z hlediska

(sobeckých) individuů, bez ohledu na to, jak by mohly být výhodné pro celou populaci či druh (Flegr, 2015).

SBS sice nedokáže vytvářet tak spektakulární adaptace jako selekce, v evoluci, a zejména makroevoluci, má ale patrně poslední slovo. Lidský mozek, například, je patrně jeden z nejkomplexnějších orgánů vůbec a spolu s vědomím bývá považován za jeden z „vrcholů“ selektivní evoluce. Není pochyb o tom, že právě mimořádná úroveň lidské nervové soustavy a souvisejících psychických schopností umožnila našemu druhu vytvořit mnohamiliardovou populaci, osídit prakticky celý povrch Země a vyrazit do jejího blízkého okolí. Stejně vlastnosti mohou ovšem být zároveň „danajským darem“ vedoucím k brzké extinkci (at' už vlivem katastrofického válečného konfliktu, nepodařeného biologického či fyzikálního experimentu, nebo jednoduše smrtící infekce schopné šíření jen v dostatečně husté a propojené populaci). Z hlediska SBS tak mohou být ve výhodě druhu a evoluční linie, které oplývají různými „pojistkami“ proti vzniku příliš efektivního mozku.

Selekce je oportunistická. Jejím působením se zlepšuje přizpůsobení organismů právě panujícím podmínkám, a to i za cenu snížení šance daného druhu či evoluční linie na dlouhodobé přežití. V naprostě stabilním prostředí by její význam patrně SBS zcela zastínil. V realistickém heterogenním a proměnlivém prostředí je tomu ale naopak. SBS je oportunistické daleko méně. Pokud nějaká vlastnost snižuje zdatnost svých nositelů, ale je zároveň výhodná z hlediska populace či druhu, působením SBS se patrně vytřídí takové populace a druhy, kde daná vlastnost nevratně převládla. Pravda, individuální výběr bude ve všech populacích upřednostňovat jedince, kteří danou, z hlediska individua nevýhodnou, vlastnost ztratí. Působením SBS ale v delším časovém měřítku převáží takové populace a takové druhy, které oplývají různými „pojistkami“ proti ztrátě dané vlastnosti. Působením SBS tak vznikají vlastnosti, které se ve zpětném pohledu zdají nápadně racionalní, takřka plánované, výhodné ne krátkodobě, ale z dlouhodobého hlediska, a „pojistky“, tj. vlastnosti nápadně restriktivní. Právě s vysvětlením takových adaptací měla v minulosti evoluční biologie patrně největší problémy.

4.2 Fenomény, za kterými může stát třídění z hlediska stability

Kromě některých typů altruistického chování, například hustotně závislé redukce zmnožení zmíněné výše, může SBS stát za řadou dalších fenoménů – zejména takových, jejichž vysvětlení cestou přirozeného výběru je tradičně považováno za problematické. Jedním z nich je pohlavní rozmnožování, jehož dlouhodobá stabilita a široké rozšíření navzdory celé řadě nevýhod, které svým nositelům oproti nepohlavním způsobům množení přináší, z něj dělají jednu z největších záhad evoluce (Flegr, 2005, 2010, 2013, 2015; Shcherbakov, 2012;

Toman, 2015). At' už byla jeho původní role jakákoli (viz např. Toman, 2015), posléze se jako klíčová vlastnost ukázala schopnost pohlavních druhů v populaci dlouhodobě udržovat velkou genetickou diverzitu různých alel (variant genů) umožňující sice neúplně perfektní, ale rychlé, a přitom vratné reakce populace na změny prostředí (Williams, 1975, str. 145-146, 149-154, 169; Flegr, 2010, 2013, 2015; Toman, 2015). Právě v tom může spočívat klíčová výhoda nad nepohlavními druhy, které se oportunisticky přizpůsobují panujícímu stavu a při každé další dostatečně rychlé a výrazné změně podmínek riskují vymření. Z hlediska individuálního výběru tak může být výhodné přejít k nepohlavnosti, prostřednictvím SBS ale převážily takové populace a takové druhy, které oplývají různými „pojistkami“ znemožňujícími či problematizujícími přechod k nepohlavnímu rozmnožování.

Nelze vyloučit, že SBS stojí i za širokým rozšířením procesů horizontálního genového transferu, tj. různých způsobů přenosu genů a jejich skupin mezi různými, často i nepříbuznými, evolučními liniemi (Markoš, 1995; Doolittle, 2014). Podobně jako u pohlavního rozmnožování byl původní účel horizontálního genového přenosu patrně odlišný, z hlediska celé evoluční linie ale mohou podobné „výpůjčky“ výrazně urychlit přizpůsobení panujícím podmínkám prostředí (včetně například působení antibiotik). Podobné procesy jsou z hlediska individua silně rizikové a jedinci, kteří jsou vůči nim odolní, mohli být v selekční výhodě. Působením SBS ale vznikly a rozšířily se „pojistky“ proti ztrátě schopnosti podléhat horizontálním genovým transferům, které se projevují například nápadnou stabilitou genetického kódu.

V neposlední řadě může být proces SBS zodpovědný také za fenomén evoluční stáže, respektive punktuacionalistický charakter evoluce (zejména pohlavních linií (viz např. Eldredge a Gould, 1972; Gould, 2002, str. 745-1024). Působením SBS se v průběhu existence druhu hromadí v genofondu funkčně propojené alely (varianty genů). Vzhledem ke složitosti genetické architektury dnešních organismů jsou geny ve svém vlivu na fenotyp organismu těsně a netrvále provázané. Některé z alel jsou přitom v genofondu populace udržovány frekvenčně závislou selekcí – jsou výhodné, dokud jsou vzácné, a nevýhodné, když jejich zastoupení v genofondu stoupne. Nemohou se tak jednoduše fixovat, nebo vymizet, a v genofondu druhu se hromadí (Flegr, 2010, 2013, 2015). Právě tyto alely a alely s nimi ve vlivu na fenotyp organismu těsně svázané (alely s pleiotropními účinky, tj. ovlivňující mnoho znaků, nebo zapojené do složitých sítí epistatických interakcí), postupně stabilizují složení genofondu a způsobí přechod druhu z krátké fáze evoluční plasticity, ve které podléhá klasické darwinovské evoluci, do mnohem delší fáze evoluční elasticity. Druh v elastické fázi sice může na změny prostředí reagovat rychlými (ale snadno

vratnými) změnami v zastoupení již přítomných, velmi variabilních alel (viz výše), výraznější jednosměrné a nevratné přizpůsobení je ale možné jen za speciálních podmínek – po opětovném zplastičtění (viz Flegr, 2010, 2013, 2015).

Předchozí příklady fenoménů, které může dobře vysvětlit SBS, rozhodně nejsou vyčerpávajícím seznamem. Explanatorní síla SBS se rozhodně neomezuje na oblast biologické evoluce. Notoricky známé evolučně-biologické fenomény jsme zvolili proto, že jsou nám nejbližší. V chystaném článku (Flegr a Toman, 2017) zmiňujeme také řadu fenoménů z dalších oblastí, a to včetně makroevoluce, ekologie, geofyziologie a kulturní evoluce. Jejich detailní popis by ale šel nad rámec tohoto příspěvku. Na závěr proto alespoň ve větším detailu rozebereme jeden fenomén, který se týká konkrétního průběhu makroevoluce i pokusů o její virtuální simulaci – trend zvyšování komplexity v evoluci.

4.3 Otevřená evoluce a růst organismální komplexity

Samotné fungování přirozeného výběru je do jisté míry triviální – tento princip se koneckonců využívá v nejrůznějších oborech od biotechnologií po programování. Daleko větší otázkou je, za jakých podmínek může na základě tohoto procesu docházet k otevřené evoluci. Flegr (2005) zmiňuje nejméně jednu další podmítku, kromě dědičnosti vlastností a nadprodukce variabilního potomstva, a to dostatečnou komplexitu prostředí. De Vladar, Santos a Szathmáry (2017) se problému věnují do větší hloubky, a to s přihlédnutím k pokusům o virtuální simulace otevřené evoluce. Rovněž zmiňují komplexitu prostředí, navíc se zdůrazněním role vzájemných interakcí mezi organismy. V praxi je ale zřejmě důležitý také dostatečně bohatý genetický systém, jehož prvky nekódují jen informaci, ale i způsob, jak vytvořit dceřiné entity, a to v neomezeném množství možných variací. Dále je důležité neomezené množství možných prostředí, se kterými se organismy mohou setkat (to není takový problém, protože prvkem prostředí jsou i samy měnící se organismy), a možnost pasivní či aktivní disperze alespoň některých organismů, prostřednictvím které jsou zajištěny neomezené variace prostředí všech entit v systému. De Vladar, Santos a Szathmáry (2017) dále rozlišují tři typy otevřenosti evoluce: Slabá otevřenosť umožňuje neomezený vznik nových fenotypů. Silná otevřenosť se vyznačuje průběžným vznikem výrazných evolučních novinek a inovací. Ultimátní otevřenosť umožňuje vzestup komplexity entit v systému, vznik nových úrovní organizace a nových způsobů dědičnosti.

Klasické modely evoluce nedosahují ani na úroveň slabé otevřenosťi. Některé modely umělého života po nějaký čas umožňují vznik nových fenotypů, všechny ale bez dalších zásahů po určité době spějí k ustálenému stavu. Ten se nejčastěji projevuje

maximální možnou redukcí, tj. zmenšením a zjednodušením replikujících se entit, zajišťujícím co nejrychlejší a nejefektivnější množení (viz např. Ray, 1993, 1997; Thearling a Ray, 1994, 1996; Yaeger, 1994; Ray a Hart, 1998; Sayama, 1999; Adami a kol., 2000; Channon, 2001; Suzuki a kol., 2003). Takové entity se nadále prakticky nemění. Podobně dopadly i provedené pokusy o simulaci otevřené evoluce v laboratoři (viz např. Spiegelman a kol., 1965; Oehlenschläger a Eigen, 1997). Patrně se není čemu divit, protože nesplňují ani výše uvedená bazální kritéria otevřené evoluce – komplexita prostředí, genetický systém organismů či mapování mezi „genotypem“ (zakódováním organismů) a „fenotypem“ (virtuální reprezentaci organismů ve vztahu k prostředí) jsou značně omezené. Možnosti evoluce dostatečně výrazných, a přitom s životem slučitelných změn, evolučních novinek, které vedou k výrazné změně pomyslné adaptivní krajiny a které jsou charakteristické pro silnou otevřenosť evoluce, je potom v takto jednoduchých modelech úplně vyloučena. Na zvláštním charakteru takových změn a jejich pozadí ostatně nepanuje shoda ani v oboru evoluční biologie. Jednotlivé vlastnosti organismů navíc mohou v průběhu času sehrát roli exaptací pro zcela jinou funkci. Omezený tak není ani stavový prostor evoluce a samotná evoluce není algoritmizovatelná.

Největší problém z hlediska evoluční biologie i umělého života nicméně představuje ultimátní otevřenosť evoluce, tj. neomezené zvyšování komplexity organismů v průběhu evoluce (viz Oehlenschläger a Eigen, 1997; Bedau, 2000, 2009; Channon a Damper, 2000; McMullin, 2000; Watson, 2006). Nejen, že za přirozených podmínek, tj. například bez zavedení silné umělé selekční výhody komplexnějších entit, ke zvyšování komplexity nevedla žádná dosud provedená virtuální simulace evoluce ani experimentální evoluce jednoduchých systémů v laboratoři (viz výše), ale nad existencí či podstatou tohoto trendu v živé přírodě nepanuje konsenzus ani z teoretického hlediska (viz např. McShea, 1991, 2015). U jednoduchých modelů zvlášť chybí charakteristické spojování do jisté míry rigidních modulů do entit vyšší hierarchické úrovně typické pro biologickou evoluci, která trend ke zvyšování maximální úrovně této hierarchické komplexity patrně opravdu vykazuje (viz např. McShea, 1991, 2015; Bedau a kol., 2000; Watson, 2006).

Potenciál ke zvyšování hierarchické úrovně v evoluci a reálné naplnění tohoto potenciálu jsou ovšem dvě odlišné věci. Právě zde do hry vstupuje SBS, které mohlo sehrát roli rozhodujícího faktoru vedoucího k naplnění tohoto potenciálu. Prostřednictvím SBS se totiž v průběhu času hromadí skupiny funkčně provázaných alel, které patrně nemohou dojít fixaci či eliminaci za žádných realistických podmínek (viz výše). Působením SBS rovněž zákonitě dochází k akumulaci takových skupin

funkčně provázaných genů, které již nemohou odpovídat na selekční tlaky (může se jednat například o znaky kódované mnoha vzájemně zastupitelnými geny nebo naopak znaky kódované geny, jejichž sebemenší změna by vedla k neživotaschopnosti organismu). Jak si všímá teorie zamrzlé evoluce (Flegr, 2010, 2013, 2015), takové znaky a jejich provázané skupiny v evoluci evoluční linie jednosměrně přibývají a snižují její evolvabilitu, schopnost podléhat adaptivní evoluci. Jedná se navíc patrně o proces připomínající Mullerovu rohatku, tj. nevratný proces hromadění slabě škodlivých mutací v liniích pohlavních organismů. Pomalé hromadění trvale stabilních znaků probíhá po malých krůčcích ve všech pohlavních liniích, takže jej nedokáže zastavit ani druhový výběr zvýhodňující linie s nejvyšší zbývající evolvabilitou.

Cestou z této jednosměrné pasti je zvýšení hierarchické úrovně tělní organizace. To může nastat modularizací stávající tělní organizace, nebo spojením několika příbuzných či nepříbuzných entit do modulárního celku vyšší úrovně a správou jednotlivých modulů jako oddělených a jen omezeně proměnlivých celků. Typickým případem prvního děje je například vznik ontogeneze mnohobuněčných živočichů. Splývání stejných celků potom nastává například při vzniku kolonií a mnohobuněčnosti, splývání různých entit zase při různých formách symbiozy a symbiogeneze. V další evoluci může docházet k rozrůzňování modulů, jejich odlišné regulaci, jejich fúzim, duplikacím, nasazení na jiných místech apod. Modulární charakter se působením SBS na vyšší úrovni stírá a celý proces se po nějaké době opakuje nanovo (Flegr, 2015).

K uvedeným případům zvýšení úrovně hierarchické organizace sice může docházet i bez nevratného „zamrzání“ znaků, pouze tento proces ale vytváří komplexní tlak čím dál více zvýhodňující zpočátku třeba i suboptimálně fungující entity vyšší úrovně s do značné míry obnovenou evolvabilitou. Postupnému „zamrzání“ znaků a snižování evolvability nasvědčuje řada genetických, ekologických a paleobiologických dokladů, které shrnuje např. Flegr (2015). Teorie zamrzlé evoluce také dokáže vysvětlit řadu záhadných doprovodných fenoménů, jako například monotónní nárůst maxima hierarchické komplexity v evoluci, postupnou akceleraci trendu, skoková zrychlení trendu v pozdním neoproterozoiku a kambru, kdy se začala rozvíjet primárně pohlavní eukaryota a mnohobuněční živočichové, u kterých teorie předpokládá nejvýraznější míru nevratného „zamrzání“, diverzifikaci modulů na bezprostředně nižší úrovni tělní organizace a ubývání diferenciace na stupních nižších. Ač se to může jevit jako kontraintuitivní, je možné, že k replikaci podobného trendu v simulacích umělého života a virtuální evoluce bude třeba zavést podobný, z SBS vyplývající a všeobecný faktor postupného snižování evolvability.

5 Závěr

Přírodní výběr není jediný, ale ani nejobecnější proces ovlivňující biologickou evoluci. Selekce je manifestací obecnějšího mechanismu „přežití stabilnějšího“, jehož konsekvence v systémech podléhajících evoluci jsme navrhli nazvat třídění z hlediska stability. Tento proces, jehož role v evoluci byla donedávna silně nedoceňovaná, může ve skutečnosti objasnit vznik a stabilitu řady záhadných a prostřednictvím přirozeného výběru těžko vysvětlitelných mikroevolučních, makroevolučních, ekologických a geofyziologických fenoménů, či jevů náležejících do sféry kulturní evoluce. Některé z nich, jako široké rozšíření a stabilitu (určitých forem) altruistického chování, pohlavnosti, horizontálního genového transferu a evoluční stáže, jsme popsali ve větším detailu.

Třídění z hlediska stability se uplatňuje na všech úrovních ve všech systémech podléhajících evoluci, tj. s historií, respektive pamětí. Všechny znaky organismů fixované v průběhu biologické evoluce mají zřejmě původ ve spolupůsobení třídění z hlediska dynamické stability, tj. přirozeného výběru, a třídění z hlediska statické stability – sil, které posouvají systém ke statické, respektive dynamické, stabilitě – v dostatečně komplexním prostředí. Podobné rysy patrně musí vykazovat všechny dostatečně komplexní přirozené i umělé systémy, ve kterých v závislosti na sobě vznikají vzájemně si konkurenční entity vykazující dědičnost a variabilitu – například kulturní evoluce nebo různé virtuální simulace biologické evoluce. Třídění z hlediska stability tak nabízí vhled do evoluce jakéhokoli komplexního systému.

Poděkování

Tento příspěvek vznikl s podporou Grantové agentury Univerzity Karlovy v rámci projektu GAUK č. 578416 (Modularita a její role v biologické evoluci) a univerzitního výzkumného centra UK UNCE 204004.

Literatura

- Adami, C., Ofria, C., Collier, T. (2000). Evolution of biological complexity. *Proceedings of the National Academy of Sciences of the United States of America*. 97(9): 4463-4468.
- Bardeen, M. (2009). Lessons from Daisyworld. Survival of the stable. Disertační práce, Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton.
- Bardeen, M., Cerpa, N. (2015). Editorial: Technological evolution in society – The evolution of mobile devices. *Journal of Theoretical and Applied Electronic Commerce Research*. 10(1): 1-7.

- Bedau, M. (2009). The evolution of complexity. V *Mapping the Future of Biology: Evolving Concepts and Theories*, str. 111-130. Springer.
- Bedau, M., McCaskill, J., Packard, N., Rasmussen, S., Adami, C., Green, D., Ikegami, T., Kaneko, K., Ray, T. (2000). Open problems in artificial life. *Artificial Life*. 6(4): 363-376.
- Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. *Philosophy of Science*. 75(5): 560-570.
- Bouchard, F. (2011). Darwinism without populations: a more inclusive understanding of the „Survival of the Fittest“. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. 42(1): 106-114.
- Bourrat, P. (2014). From survivors to replicators: evolution by natural selection revisited. *Biology and Philosophy*. 29(4): 517-538.
- Channon, A. (2001). Passing the ALife test: Activity statistics classify evolution in Geb as unbounded. Ve zborníku *Advances in Artificial Life: 6th European Conference, ECAL 2001 Prague, Czech Republic, September 10–14, 2001*, str. 417-426. Springer.
- Channon, A., Damper, R. (2000). Towards the evolutionary emergence of increasingly complex advantageous behaviours. *International Journal of Systems Science*. 31(7): 843-860.
- Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. John Murray.
- Darwin, C., Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties a species by natural means of selection. *Journal of the proceedings of the Linnean Society of London. Zoology*. 3(9): 45-62.
- Dawkins, R. (2003). *Sobecký gen*. Mladá fronta, Praha.
- de Vladar, H., Santos, M., Szathmáry, E. (2017). Grand views of evolution. *Trends in Ecology & Evolution*. V tisku.
- Doolittle, W. (2014). Natural selection through survival alone, and the possibility of Gaia. *Biology and Philosophy*. 29(3): 415-423.
- Eldredge, N., Gould, S. (1972). Punctuated equilibria: an alternative to phyletic gradualism. Ve zborníku *Models in Paleobiology*. Freeman, Cooper and Co.
- Flegr, J. (2005). *Evoluční biologie*. Academia, Praha.
- Flegr, J. (2010). Elastic, not plastic species: frozen plasticity theory and the origin of adaptive evolution in sexually reproducing organisms. *Biology Direct*. 5.
- Flegr, J. (2013). Microevolutionary, macroevolutionary, ecological and taxonomical implications of punctuational theories of adaptive evolution. *Biology Direct*. 8.
- Flegr, J. (2015). *Evoluční tání*. Academia, Praha.
- Flegr, J., Toman, J. (2017). Stability-based sorting: The forgotten process behind (not only) biological evolution. *Journal of Theoretical Biology*. Zasláno do tisku.
- Godfrey-Smith, B. (2009). *Darwinian populations and natural selection*. Oxford University Press.
- Gould, S. (2002). *The structure of evolutionary Theory*. The Belknap Press of Harvard University Press.
- Hladký, V., Kočandrle, R. a Kratochvíl, Z. (2012). *Evoluce před Darwinem*. Pavel Mervart.
- Komárek, S. (2004). *Mimikry, aposematismus a příbuzné jevy: Mimeticismus v přírodě a vývoj jeho poznání*. Dokořán.
- Komárek, S. (2008). *Příroda a kultura: Svět jevů a interpretací*. Academia.
- Laland, K., Uller, T., Fellman, M., Sterelny, K., Muller, G., Moczek, A., Jablonka, E., Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B-Biological Sciences*. 282(1813): 1-14.
- Lotka, A. (1922). Natural selection as a physical principle. *Proceedings of the National Academy of Sciences of the United States of America*. 8(6): 151–154.
- Markoš, A. (1995). The ontogeny of Gaia: the role of microorganisms in planetary information network. *Journal of Theoretical Biology*. 176(1): 175-180.
- Matthew, P. (1831). *On naval timber and arboriculture: with critical notes on authors who have recently treated the subject of planting*. Adam Black; Longman, Rees, Orme, Brown, and Green.
- Maynard Smith, J., Price, G. (1973). *The logic of animal conflict*. Nature. 263: 15-18.
- Maynard Smith, J., Szathmáry, E. (2010). *The major transitions in evolution*. Oxford University Press.

- Mayr, E. (2003). *The growth of biological thought: Diversity, evolution, and inheritance*. The Belknap Press of Harvard University Press.
- McMullin, B. (2000). John von Neumann and the evolutionary growth of complexity: Looking backward, looking forward. *Artificial Life*. 6(4): 347-361.
- McShea, D. (1991). Complexity and evolution: What everybody knows. *Biology and Philosophy*. 6(3): 303-324.
- McShea, D. (2015). Three trends in the history of life: An evolutionary syndrome. *Evolutionary Biology*. 43(4): 1-12.
- Oehlenschläger, F., Eigen, M. (1997). 30 years later – a new approach to Sol Spiegelman's and Leslie Orgel's in vitro evolutionary studies - Dedicated to Leslie Orgel on the occasion of his 70th birthday. *Origins of Life and Evolution of the Biosphere*. 27(5-6): 437-457.
- Okasha, S. (2006). *Evolution and the levels of selection*. Oxford University Press.
- Pascal, R., Pross, A. (2015). Stability and its manifestation in the chemical and biological worlds. *Chemical Communications*. 51(90): 16160-16165.
- Pross, A. (2012). *What is life? How chemistry becomes biology*. Oxford University Press.
- Rappaport, R. (1999). *Ritual and religion in the making of humanity*. Cambridge University Press.
- Ray, T. (1993). An evolutionary approach to synthetic biology: Zen and the art of creating life. *Artificial Life*. 1(1-2): 179-209.
- Ray, T. (1997). Evolving complexity. *Artificial Life and Robotics*. 1(1): 21-26.
- Ray, T., Hart, J. (1998). Evolution of differentiated multi-threaded digital organisms. Ve zborníku *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*. MIT Press.
- Sayama, H. (1999). A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life*. 5(4): 343-365.
- Shcherbakov, V. (2012). Stasis is an inevitable consequence of every successful evolution. *Biosemiotics*. 5(2): 227-245.
- Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*. 106(6): 467-482.
- Slobodkin, L., Rapoport, A. (1974). An optimal strategy of evolution. *The Quarterly Review of Biology*. 49(3): 181-200.
- Spiegelman, S., Haruna, I., Holland, I., Beaudreau, G., Mills, D. (1965). The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proceedings of the National Academy of Sciences of the United States of America*. 54(3): 919-927.
- Suzuki, H., Ono, N., Yuta, K. (2003). Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life*. 9(2): 153-174.
- Toman, J. (2015). *Pohlavní rozmněžování optikou evoluce: Vznik vývoj a paradoxy největší evoluční záhad*. Academia.
- Thearling, K., Ray, T. (1994). Evolving multi-cellular artificial life. Ve zborníku *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. MIT Press.
- Thearling, K., Ray, T. (1996). Evolving parallel computation. *Complex Systems*. 10(3): 229-237.
- Van Valen, L. (1989). Three paradigms of evolution. *Evolutionary Theory*. 9(2): 1-17.
- Wagner, N., Pross, A. (2011). The nature of stability in replicating systems. *Entropy*. 13(2): 518-527.
- Watson, R. (2006). *Compositional evolution: The impact of sex, symbiosis, and modularity on the gradualist framework of evolution*. MIT Press.
- Williams, G. C. (1975). *Sex and evolution*. Princeton University Press.
- Wynne-Edwards, V. (1986). *Evolution Through Group Selection*. Blackwell Scientific Publications.
- Yaeger, L. (1994). Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or Polyworld: Life in a new context. Ve zborníku *Artificial Life 3: Proceedings of the Third International Conference on the Synthesis and Simulation of Living Systems*. Addison-Wesley.

Effect of stimulus distribution on the buildup of contextual plasticity in sound localization

Beáta Tomoriová, Gabriela Andrejková, Norbert Kopčo

Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University in Košice
beata.tomoriova@upjs.sk, gabriela.andrejkova@upjs.sk, norbert.kopco@upjs.sk

Abstract

Contextual plasticity (CP) is a short-term adaptation effect in sound localization (Kopco et al., 2007). CP is exhibited by biases in localization of click stimuli in target-alone trials when, on interleaved contextual trials, a distractor-target click pair is presented. In previous CP studies, a gradual drift in localization responses towards the front was observed. The drift was not attributable to CP but can potentially confound its analysis. Here, we analyze several CP data sets with various distributions of contextual and target-alone stimuli in order to test whether the drift is related to asymmetry in the distribution of the stimuli relative to the subject's midline. The results show that the drift tends to have a steeper slope when the stimulus distribution mean is more lateral, confirming that it is likely related to the distribution asymmetry.

1 Introduction

The perceived location of a sound source can be affected by sounds heard shortly prior to the sound of interest. One example is the precedence effect (for review, see Litovsky et al., 1999), in which presenting an identical sound a few milliseconds ahead of the target induces a percept of only one fused sound coming from the direction of the first of the sounds. Another example is when the target sound is preceded by a longer-lasting "adaptor" sound, resulting in localization shifts away from the adaptor location (e.g., Carlile et al., 2001; Kashino & Nishida, 1998).

Here, we examine another form of adaptation, referred to as *contextual plasticity* (CP), in which the perceived location of a target is affected not only by a single preceding stimulus but by stimuli presented within a broader temporal window. CP was first observed in a study by Kopco et al. (2007). In that study, trials in which only the target stimulus was presented were randomly interleaved with trials in which the target stimulus was preceded by an identical distractor stimulus (presented from a fixed location, either frontal or lateral). Unexpectedly, responses on target-alone trials were found to be affected by the location of the distractor in the interleaved trials, being more frontal when distractor was lateral and vice versa

(i.e., the responses were shifted away from the distractor location). The induction of contextual biases was confirmed in several other CP experiments in which CP was evaluated relative to a baseline condition only containing target-alone trials (Andrejkova et al., 2016; Tomoriová et al., 2014; Kopco et al., 2015).

However, the previous CP studies also revealed that responses, including those from a reference baseline condition with no distractor, tend to gradually drift towards the listener's straight ahead (azimuth 0°). The cause of these drifts is not yet understood and its examination is important for the CP analyses, e.g., to clarify to what extent gradual changes in responses relative to baseline responses correspond to buildup/decay of CP and to what extent they are altered by the (potentially also context-dependent) drift.

In the current study, we analyze datasets from several previous CP experiments in order to examine the drifts for different context conditions. Based on preliminary analyses (described in more detail in subsection 3.2 Preliminary analyses below) we hypothesize that the drifts are related to the fact that the sound stimuli are not distributed around straight ahead where the localization acuity is the highest (Makous & Middlebrooks, 1990), but at the side. Assuming that the spatial distribution of incoming sounds is approximately uniform (perhaps with stimuli coming from the median plane being more likely since we tend to face the stimulus of interest), and that the spatial representation of the stimulus acoustic cues (ITD and ILD) covers the range of -90° to +90°, the mean of the distribution of the stimuli can be *a priori* expected to be 0°. Then, in the current experiments in which stimuli are presented from one quadrant (e.g., from 0° to 90° with mean at around 45°), it is likely that the spatial auditory representation adapts to map this limited range to the *a priori* expected range of -90° to 90°, thus causing a drift in the perceived location of the stimuli. Such adaptation has been hypothesized for auditory (Maddox et al., 2014) and visual (Grossberg, 1982) spatial representations, resulting in improved spatial resolution, and it has been demonstrated for ILD representation (Dahmen et al., 2010). Based on this hypothesis we expect that the more lateral is the bias in distributions of stimuli, the larger drift in temporal data

towards midline will be observed. In this analysis we assume that each type of stimulus (distractor, target from a target-alone trial, target from a context trial) contributes equally to this effect since the stimuli are identical in their acoustic properties.

2 Methods

We performed analyses on data from four previous CP experiments. The experiments are described here only briefly, more details can be found in following publications: Exp1: Kopco et al. (2007), Exp2: Andrejkova et al. (2016), Exp3&4: Tomoriova et al. (2014). (Note: one experimental condition from Exp 4, with stimuli centered around azimuth 90 degrees was omitted from analyses in this paper, because of possible confounds related to front-back confusion errors - see, e.g., Makous and Middlebrooks, 1990).

2.1 Previous experiments

In each experiment, the subject was seated in a classroom or anechoic room (Exp 1&2) or acoustic booth (Exp 3&4) and was surrounded by 9 loudspeakers arranged uniformly in a quarter-circle arc. The orientation of the listener relative to the speaker array, as well as the distribution of stimuli across the speaker array, varied across different experimental conditions and experiments (see Fig. 1 for an example of setup used in Exp 3, and Fig. 2 for distributions of stimuli in each condition of each experiment). Symmetric conditions for left and right orientation of speaker array relative to the listener were tested but in all presented analyses data are collapsed across orientation.

The subject's task was to point with a hand-held pointer to a perceived location of a target sound. *Experimental runs* contained a mixture of "target-alone" trials in which only the target was presented and "contextual trials" in which the target was preceded by an identical distractor. Additionally, Exp 3&4 contained reference *baseline runs* consisting of target-alone trials only. Experimental runs in Exp 3&4 started and ended with several pre-adaptation and post-adaptation trials intended to examine the buildup/decay of adaptation after onset/offset of the contextual trials.

Two-ms frozen noise bursts ("click" stimuli) were used as target stimuli in all experiments. In a majority of the experiments, the distractor was identical to the target. However, in a subset of contextual trials of Exp 2 the 1-click distractor was replaced by an 8-click distractor. Table 1 along with Fig 2. summarizes basic information about the experiments.

	baseline condition	distractor type
Exp 1	not included	1-click
Exp 2	not included	1-click / 8-click
Exp 3	included	1-click
Exp 4	included	1-click

Tab. 1: Summary information about the differences in design and stimuli across the experiments. Differences in distributions of stimuli are summarized in Fig. 2.

2.2 Analyses of the drift

In order to evaluate how slope of the drift depends on the laterality of the distribution we computed the mean stimulus lateral position in degrees (*mp* value in Fig. 2) for each condition as the weighted mean of stimulus locations used in that condition (each location was weighted by the number of stimuli presented from that location), and fitted the temporal profile data for each subject and condition (across-subject average of these data is shown in Fig. 2) by the linear function:

$$y = k * x + q$$

in which *y* corresponds to an estimate of the response bias relative to the actual target location (i.e., localization error) averaged across all target locations, *x* corresponds to a subrun number (trials were grouped into subruns in this analysis; one subrun corresponds to 42 trials in Exp 1, 35 trials in Exp. 2 and 28 trials in Exp 3&4) within the adaptation part of the run, and parameters *k* and *q* represent the slope of the temporal drift and its offset, respectively.

Only the responses from the target-alone trials are considered in the analyses of the drifts.

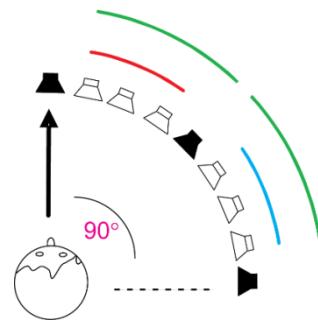


Fig. 1: Experimental setup for the Exp. 3. Black loudspeakers indicate possible distractor locations (only one distractor location was used in each experimental run), while target could be presented from each of the 7 inner loudspeakers (in target-alone-trial targets) or was restricted to specific subregion of the speaker array (in contextual trials) in left-half/right-half/full-context condition (coded here and in other Exp 3&4 figures by red, blue and green color , respectively).

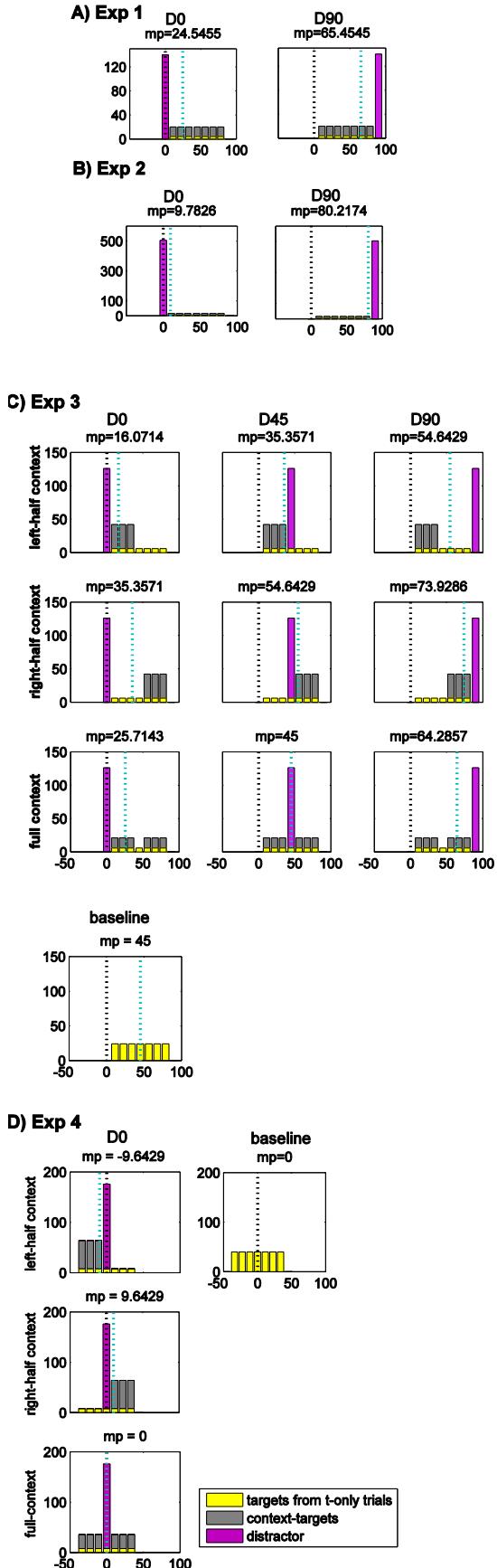


Fig. 2: Distribution of stimuli (as the number of stimuli per location) in different experimental conditions. X-axis shows stimulus lateral angles with 0° corresponding to straight ahead (depicted by black dotted line). Blue dotted line indicates the mean stimulus lateral position (mp). Proportion of different types of stimuli is depicted in each bar by different color (distractor: pink, target-alone-trial target: yellow, context-target: grey). For the 8-click distractor used in Exp 2, each click is considered as a separate stimulus.

3 Results and discussion

3.1 Distributions

The distributions of stimuli in each experimental condition (with pre- and post-adaptation trials excluded) from the experiments are plotted in Fig. 2. Different colors correspond to different types of stimuli (pink: distractor, grey: context-targets, yellow: target-alone-trial targets). Mean lateral stimulus position value computed from a distribution of stimuli is depicted by dotted blue line, while straight ahead position is depicted by dotted black line. Conditions identified by this analysis as similar in terms of their similar mean stimulus lateral position value are D0 condition from Exp 1 and D0 full-context condition from Exp 3, and similarly for D90. In Exp 3, identical mp value was found for two conditions with largely different number of stimuli per identical location, specifically, D0: right-half context vs D45:left-half context, and D90 left-half context vs D45: right-half context.

3.2 Preliminary analyses: Temporal profile of adaptation during a run

Analyses in this section are provided to illustrate a potential confounding effect of the drift on CP analysis and explain in more detail the motivation to examine drifts as a function of laterality of the distribution.

Fig. 3 shows biases in subjects' responses during the course of the run, averaged across the target locations. For both Exp 1&2, results are plotted separately for each acoustic environment (classroom or anechoic room). In both environments and both experiments, when the distractor was frontal (solid lines), responses had approximately constant or slightly increasing bias of approx. 8 to 10 degrees away from the distractor. When the distractor was lateral (dash-dotted lines), responses were initially biased by up to approx. 4 degrees but during experimental run they drifted away from the distractor location (towards subject's straight ahead). The drifts were more pronounced for Exp 2 where 8-click distractor was used on some contextual trials instead of 1-click distractor.

This pattern of results might indicate that 1) for 8-click distractor the buildup of CP lasts longer and

saturates at larger magnitudes, and 2) lateral distractor induces larger CP. However, data from Exp 3&4 indicate a different explanation of this effect. Conditions comparable to those from Exp 1 in their distribution of stimuli are Exp3 D0 and D90 full-context conditions (green lines in first and third panel; see also Fig. 2 for the corresponding distributions). Despite minor differences in initial magnitudes after the start of the adaptation part (white region), similarly to previous two experiments, frontal distractor responses are approximately constant while lateral distractor responses gradually drift towards the front. However, the drift can be observed also in the remaining conditions, and, in particular, in the baseline condition (orange line), indicating that the drift cannot be explained solely by the presence of the distractor (i.e., solely by CP effect).

But in conditions in which distractor is present, an increase in laterality of the distractor tends to increase the negative value drift (compare drifts across different columns of Exp 3) and dash-dotted vs solid lines in Exp 1&2. In the following section, we evaluate this effect.

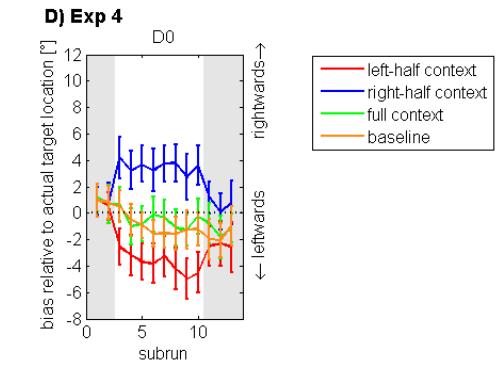
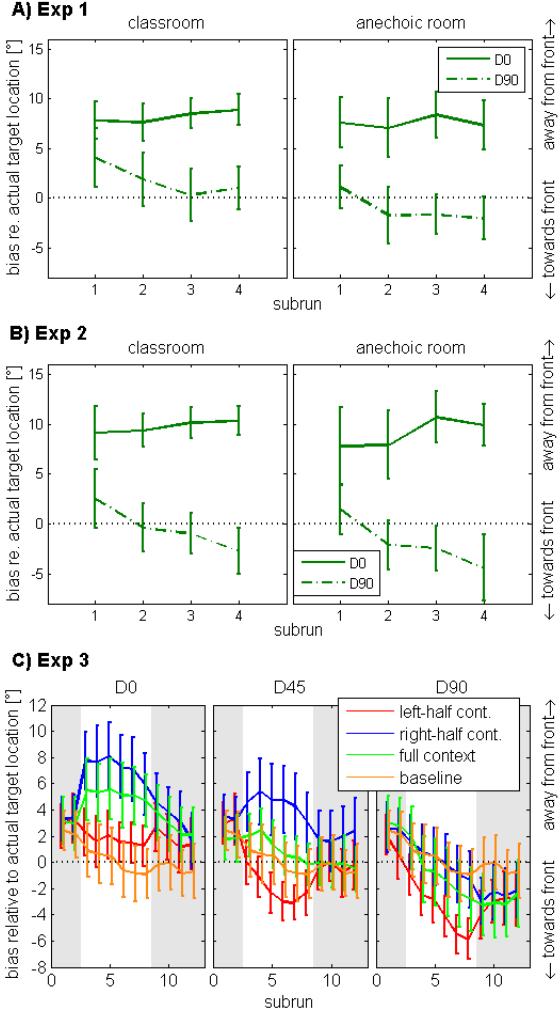


Fig. 3: Temporal profile of adaptation for different experimental conditions. Across-subject mean (+ standard error of the mean) bias relative to actual target location, averaged across all target locations, is plotted as a function of subrun. D0, D45 and D90 correspond to different distractor locations while lines of different colors correspond to conditions with different context-targets locations (see Fig. 2 for the distributions). Shaded regions indicate pre- and post-adaptation part of the run in which only target-alone trials were presented, white region indicates adaptation part in which target-alone trials were interleaved with contextual trials.

3.3 Effect of mean stimulus lateral position on slope of the response drift

Fig. 4 shows across-subject mean slope for different mp values, i.e., how slope of the drift changes with increasing distance between the mean of the actual distribution and the position straight ahead.

In Exp 1&2, k values range from approx. $-1.5^\circ/\text{subrun}$ for mp values farther from straight ahead (conditions with lateral distractor, depicted by right-pointing triangle symbols) to approx. $0.5^\circ/\text{subrun}$ for mp values closer to straight ahead (conditions with frontal distractor, depicted by left-pointing triangle symbols). Datapoints from the same experiment and acoustic environment, i.e., which are the most comparable, are interconnected. In all of them, k value becomes more negative with increasing mean stimulus lateral position. (note: mp values between the two distractor conditions in Exp 2 are more distant from each other due to the fact that 8-click distractor was used on some trials, and it shifted the mean of the distribution towards its location).

Interesting observations are a slightly positive k values for frontal distractor conditions. Their potential cause is the interference of the non-CP-related drift with longer CP buildup process, since the direction of the drift is consistent with CP buildup effect, i.e., away from the distractor.

In Exp 3&4, k values range from $-1^\circ/\text{subrun}$ to $0^\circ/\text{subrun}$. Even though the trends are similar to Exp 1&2, these results also show that the effect occurs only between conditions which differ only in distractor

location, with their context-targets locations being identical, not vice versa, where the slope value for some distractor locations even increases to less negative values (compare the datapoints plotted with different symbols of the same color versus those plotted with different colors but same symbol; e.g., compare a change in the slope between the three points of a green line vs change in the slope between the three diamonds). A different pattern for the blue line caused by the D0 right-half context condition (left-most blue triangle) might be related to several aspects of the data, e.g., the fact that for this condition, the biases are the largest in the lateral direction, thus requiring the most effortful response, or the fact that the target distribution is bi-modal (Fig 2).

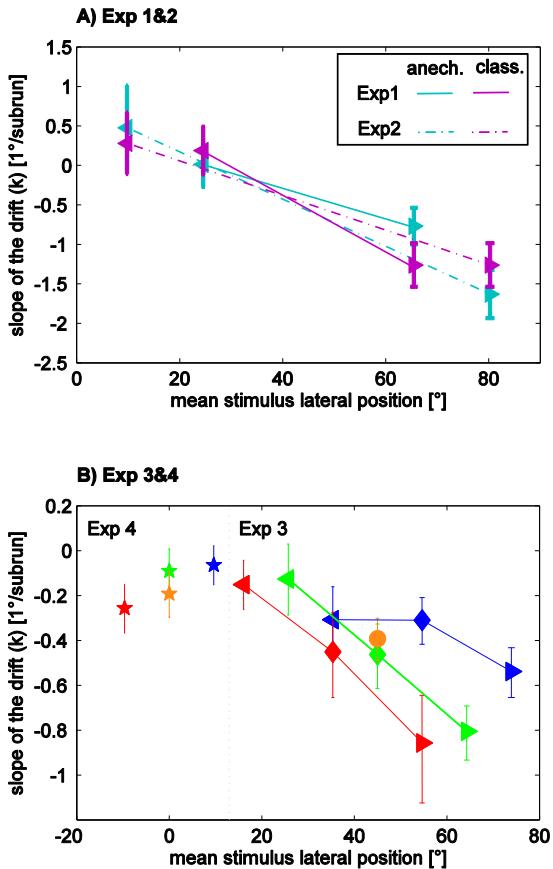


Fig. 4: Across-subject mean and standard error of the slope of the linear fit (from the arctangented parameter k , then transformed back into k values in the plot), as a function of mean lateral stimulus location. Datapoints from conditions which differ only in distractor location are interconnected for better visualization. Distractor location is indicated by the marker symbol, with left-/right-pointing triangle corresponding to D0/D90 condition and diamond symbol corresponding to D45 condition. Stars symbol also correspond to D0 condition but with different orientation of speaker array relative to subject.

For the conditions having the same mp value, k values do not differ in two cases (compare green and orange symbols at mp values of 0° and at 45°) but large difference is observed in one case (compare blue diamond and red right-pointing triangle at mp value of around 55° , corresponding to D45 right-half context and D90 left-half context, respectively). The cause of this difference is not clear, but possibly can be explained based on CP effects, in the first case acting in the opposite direction as non-CP-related drift in responses (since CP in general acts in the direction away from distractor towards the side with context-targets, see Fig. 4), decreasing the slope to less negative values and in second case acting in the same direction, increasing the negativity of the slope.

To quantify these observations, Pearson's correlation coefficient was computed between the mp values and the corresponding across-subject mean k values within each experiment. The results are summarized in Table 2. A significant correlation was found in all experiments in which it was expected, suggesting that the hypothesis is correct, even though it doesn't explain all aspects of the data.

	r	p
Exp 1&2 merged	-0.7341	0.0381
Exp 3	-0.7306	0.0164
Exp 4	0.8777	0.1223
Exp 3&4 merged	-0.7598	0.0016

Tab. 2: Pearson's correlation coefficients and their respective p-value between the mp and mean k value (from Fig. 4) for each experiment or for the combination of experiments (if comparable).

4 Summary and general discussion

Previous analyses of temporal profile of localization responses during runs revealed that responses drift towards the front, also in the baseline condition in which no distractor was presented. Here, we analyzed these drifts, hypothesizing that they emerge from the fact that distribution of stimuli during an experimental run was offset from straight ahead where spatial acuity is the largest (Makous & Middlebrooks, 1990) and auditory system might try to optimally align the stimuli with neural representation in order to better discriminate between spatial locations of presented sounds. The results are in general consistent with this hypothesis, in a sense that responses drifted towards straight ahead and, on average, the data drifted more when distribution of stimuli were centered more laterally. (Note that the slightly increasing slope in Exp 1&2 for the frontal distractor was only negligible and might be caused by CP, since it has the same direction, i.e., away from the distractor. Similar explanation can be used for Exp 4 condition in which mean lateral stimulus position was slightly offset to the left of

straight ahead but responses drifted away from straight ahead). However, several properties of the drifts were observed which need more detailed examination. First, when distractor location was kept constant and distributions differed only in context-targets locations, the increasing effect of mean stimulus laterality on the drift magnitude was not observed and in some conditions was even reversed (the slope of the drift became less negative for more lateral mean stimulus positions), but the change was on average smaller and less consistent compared to when context-targets locations were kept constant and location of the distractor varied. A possible explanation for this reversed effect is that the context-targets are always presented shortly (25–400 ms) after the distractor, i.e., the two sounds might perceptually interact (see Litovsky et al., 1999) and thus their effect on the overall stimulus distribution might be reversed or suppressed).

Another suggestion for further analysis, which would better separate CP effect from the aligning effect is to omit first adaptation subrun, during which CP effect might still build up. Alternatively, a new experiment can be designed in which localization will be examined for normal or uniform spatial distributions with no specific temporal patterns as in CP experiments. Finally, another partial contribution to the observed drifts might originate from the method of responding used in these experiments (i.e., pointing) which might cause that subjects gradually underestimate more lateral locations since pointing there requires more effort.

Acknowledgement

This contribution was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex Environments) and by the TECHNICOM project, ITMS: 26220220182, of the EU RDP.

References

- Andrejkova, G., Best, V., Shinn-Cunningham, B. G., Kopco, N. (2015). Streaming and sound localization with a preceding distractor. *The Journal of the Acoustical Society of America*, 137(4), 2228–2228.
- Carlile, S., Hyams, S., & Delaney, S. (2001). Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *The Journal of the Acoustical Society of America*, 110(1), 416–424.
- Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space. *Neuron*, 66(6), 937–948.
- Grossberg, S. (1982). *Studies of Mind and Brain*. Springer Netherlands.
- Kashino, M., & Nishida, S. (1998). Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *The Journal of the Acoustical Society of America*, 103(6), 3597–3604.
- Kopčo, N., Marcinek, L., Tomoriová, B., & Hládek, L. (2015). Contextual plasticity, top-down, and non-auditory factors in sound localization with a distractor. *The Journal of the Acoustical Society of America*, 137(4), EL281–287.
- Kopčo, N., Best, V., & Shinn-Cunningham, B. G. (2007). Sound localization with a preceding distractor. *The Journal of the Acoustical Society of America*, 121(1), 420–432.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, 106(4 Pt 1), 1633–1654.
- Maddox, R. K., Pospisil, D. A., Stecker, G. C., & Lee, A. K. C. (2014). Directing Eye Gaze Enhances Auditory Spatial Cue Discrimination. *Current Biology*, 24(7), 748–752.
- Makous, J. C., Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, 87(5), 2188–200.
- Tomoriova, B., Marcinek, L., Hladek, L., & Kopco, N. (2014). Contextual plasticity in sound localization: characterization of spatial properties and neural locus. In: *Proceedings of 7th Forum Acusticum 2014*.

Visual cognitive styles in virtual environments: Constructing and evaluating adequate tasks

Pavel Ugwitz, Vojtěch Juřík, Čeněk Šašinka

Department of Psychology, Faculty of Arts MU

HUME Lab, FF MU

Arna Nováka 1

602 00 Brno

172577@mail.muni.cz

jurik.vojtech@mail.muni.cz

cenek.sasinka@mail.muni.cz

Abstrakt

Vizuální kognitivní styly se manifestují způsoby, kterými lidé implicitně zpracovávají vizuální podněty z okolí. Je však otázkou, jak vizuální kognitivní styly měřit. Klasické způsoby měření využívají dotazníků, či jednoduchých obrázkových úloh. Tyto způsoby však ze své podstaty nemohou zachytit komplexitu skutečných vizuálních kompozicí. S nástupem (mobilních) eye-trackingových zařízení se sice možnosti měření kognitivních stylů posouvají, jedná se zde však pouze o receptive zpracování reálného vizuálního podkladu. Cílem tohoto příspěvku je představit možnosti i úskalí, které v sobě skýtá konstrukce virtuálních prostředí a jejich následné nasazení v úlohách zaměřených na zachycení vizuálních kognitivních stylů, v rámci virtuální reality, při užití eye-trackingu. Konstrukce vlastních virtuálních prostředí sice otevírá nové možnosti, technologické limitace i specifika vizuálního vnímání s sebou nicméně nejsou rádu omezení – je tedy třeba klást důraz na metodické podchycení zásad konstrukce virtuálních úloh.

Abstract

Visual cognitive styles manifest in ways humans implicitly process visual cues from their surroundings. A question arises: how are cognitive styles to be measured? Classic approaches make use of questionnaires or figural tasks. However, the nature of these approaches fails at capturing complexity of real visual compositions. With the onset of (mobile) eye-tracking devices, this limitation can be overcome; the paradigm, however, still revolves around receptive processing of existing visual material. The aim of this paper is to consider constructing 3D virtual scenes, followed by utilizing such scenes in visual cognitive styles tasks – using virtual reality and eye-tracking devices. While constructing virtual environments does indeed offer new possibilities, technological constraints and visual perception limitations do, however, introduce some new limitations to the table. Therefore, offering a methodological grasp of virtual scene construction is relevant.

1 Introduction to visual cognitive styles

As characterized by some of the pioneers in the field of cognitive differences research (Ausburn & Ausburn, 1978), cognitive style is: *a psychological dimension that represents consistencies in how an individual acquires and processes information*. A visual cognitive style is a further, finer differentiation of the former concept. However, since this metaphorical terrain of cognitive styles research has grown ever so complex over the years (Kozhevnikov, 2007), it is also worth mentioning that in this proposal, we chose to adhere to the definition of visual cognitive styles formulated by Kozhevnikov & colleagues (Kozhevnikov et al., 2005).

According to Kozhevnikov's classification, there are categories of object visualizers and spatial visualizers. Their definition is as follows ("Object-Spatial-Verbal Cognitive Style Model"):

- **Object visualizers** prefer to construct vivid, concrete and detailed images of individual objects; they rely primarily on visual-object strategies; they do better on object imagery tasks.
- **Spatial visualizers** prefer to schematically represent spatial relations of objects and spatial transformations; they rely primarily on visual-spatial strategies; they do better on spatial imagery tasks.

To measure visual cognitive styles, the original authors have been using the *Object-Spatial Imagery questionnaire* (Blajenkova et al., 2006). While taking this into consideration as a baseline, in our research, we intended to propose computerized ways of measuring visual cognitive styles that make use of eye-tracking within 3D virtual environments.

We shall review potential eye-tracking solutions that can be implemented into 3D virtual environments. Some technological constraints of these solutions and the visualization devices in use will be considered. Then, we'll delve into the problematics of constructing 3D scenes salient for eye-tracking stimuli, and go over

scene optimization (so as to ensure smoothness of data collection and future data analysis). In conclusion, we'll mention some other approaches to visual cognition testing that can be used to complement the eye-tracking method.

2 Methods: focus on eye-tracking

There is, without a doubt, a multitude of ways of measuring visual cognitive styles. These may include: *questionnaires, verbal statement analysis, map drawing tasks, path tracking, etc.* Some of these methods may be used as supplementaries to virtual environment-based testing, some of them may be even implemented into the virtual environment-based testing. However, for this paper, we have decided to focus on the VR eye-tracking method – since our context (VR headset + eye-tracker implanted into the headset) makes this a method that cannot exist beyond virtual reality.

By using a virtual reality headset with eye-tracking capability, extracting valid visual cognition data out of a real-time 3D virtual world is entirely possible. There are, however, different possibilities of collecting that eye-tracking data (and the choice of eye-tracking solution influences 3D scene creation, and vice versa). The two key solutions to eye-tracking that are to be considered are:

- **Heatmap tracking¹.** As known from simple, static, two-dimensional eye-tracking applications, this approach collects the X/Y coordinates of the eye; after the data collection phase is done, a mathematical model plots the most/least frequented gaze areas onto the 2D plane, by adding a colored “heatmap” overlay on top of the plane. Whereas this approach is relatively easily applicable and interpretable in static 2D solutions, with 3D solutions, this gets more complex, as the third dimension adds the Z axis (X/Y/Z), and a non-static movement through the 3D environment introduces an ever-changing viewport. The potential level of measurement accuracy is great; however, both technical implementation and data analysis of this approach are rather complex.
- **Object tracking.** The eye tracking happens on the level of 3D objects. Any object in the scene has two states: they are either being fixated by eye-tracking, or not. This allows for simplicity in data analysis² and implementation; on the other hand,

¹3D heatmap tracking seems to be a niche field. We are yet to find an open source solution that is freely available and usable within the bounds of the Unity graphics engine we are using. Irregardless, we have encountered a commercial 3D heatmap solution (“EyeSee3D website”).

²In terms of data output, this approach produces a series of gaze sequences, e.g. ABDEBABC, where each individual example letter represents a sequentially fixated object of the scene. This ties

when gazing into details of a single object, there is no further data discrimination. In other words, the pros & cons of the object tracking approach are the direct opposite of the heatmap approach.

Due to the aforementioned complexities of 3D dynamic heatmap tracking, we opted for the object tracking approach. This means that the elements of the scene shall be broken into individual 3D objects, so that data can be collected on said objects.

3 Reviewing the technical equipment

For any visual cognition measurements to make sense, let it be known that constructing actual 3D graphics visual scenes (along with a functional data-collecting API³) is required. Let's not focus on the technicalities of constructing such scenes here; let's, however, cover the psychological aspects of creating, and experiencing such scenes.

Every visual scene differs from the other; every single one is a unique piece – and from this perspective, one could argue that what we proposed to do here is severely lacking and rather wild (as in considering all the possible variations/interactions of variables), in terms of scientific standards. Such notion holds true indeed – up until to the point where we grasp the ongoing phenomenon of perceiving a visual display of space and objects, and subject it to more rigorous approach. Let's do that in this narrative right now; in fact, let's start from the “lens” that gateways our perception of the visual world – i.e., the visual interface.

The hardware properties of a visualization interface (a virtual reality headset⁴) allow or limit us from using it in certain ways. In essence, this can be compared to the limitations of human sight – e.g., one cannot simply perceive far-away objects in great detail, see clearly in pitch-black darkness, etc. This consideration should go double for the engineers of the VR visualization interfaces, as they should always keep in mind that not only do their devices carry their inherent limitations,

to graph theory; in-depth overview into the possibilities and data interpretation of this approach is covered in the following paper: (Dolezalova, Popelka, 2016). In short, the idea is about discovering similar (sub)patterns of eye movement people incline to – from this, sort of a cluster analysis of visual cognition patterns can be extracted.

³API stands for *Application Program Interface*. Where does this API come from? Let there be a 3D graphics engine - Unity (we are using version 5.4). It is worth mentioning that the whole of Unity is written in the C# programming language; due to the engine's massive extendability, one who is able to program in the C# language is also able to implement a system of data collection program routines. This served as our API, a way of obtaining eye-tracking data.

⁴For a recent overview of appropriate devices, i.e. virtual reality headsets, refer to a recent overview article (Lamkin, 2017). To further explain the bold claim of these devices being “appropriate”, this is so due to these devices occupying the whole of the visual *field of view* of any participant – as opposed to viewing scenes in other, less immersive visualization interfaces (such as computer screens), where objects off the screen can intervene with visual processing.

these limitations also accumulate on top the constraints of people's own, natural eyesights.

Smoothness of the experience is a necessary prerequisite of any non-negatively inhibited cognitive processes to take place in the participants. It is necessary to ensure that the perception of the participants is not, in any way, distorted in some of them (or, in consideration of the tech limitation: distorted for all the participants to the same limiting extent, and not less). This translates to technical specifications like refresh rate, or the use of high-persistence display (i.e. to experience an unblurred vision while looking around the virtual environment rendered from within the VR headset interface).

The absence of experiencing simulation sickness (Stanney et al., 2008) does play into this, too. When a man is feeling sickly (Howarth, Costello, 1997), they have a tendency of diverting their attention from cognitive/visual processing to the bodily experience created by the ill-causing stimuli and/or to preoccupying themselves with the ill-causing stimulus itself. An individual under the influence of simulation sickness can become agitated and (nearly) devoid of frustration tolerance or other compensation mechanism; this, in turn, makes them a poor subject for visual cognition testing. The fact that some individuals are more predisposed to experiencing motion/simulation sickness than others (Basu et al., 2016) calls for experimental interfaces and stimuli being useddesigned in a way that respects these highly sensitive participants.

Display acuity of the experience tells the experimenter what sorts of visual data are (not) worth planting/exploring/analyzing. Using very small objects, or objects very far away, in a visualization that does not account for the limitations of detail perceived by the human fovea or displayed by the (limited) subpixel density of a visualization device⁵, renders such tiny objects insignificant.

Optical illusions, or any sorts of dissonance-inducing stimuli (such as unbelievable vistas) are most certainly not wanted either, as such subjects have the potential of swaying the participant to investing most of their attention to such unnatural occurrences.

When supplementary devices (e.g. the integrated eye-tracker) are used, it is worth mentioning that these devices, unless wholly integrated into visualization devices in an effective 1:1 variable conversion design, introduce limitations of their own. If their differentiation (of resolution, of measurement frequency) is of greater magnitude than the one of the visual interface, a man need not worry. If, however, they differentiate at a lower rate⁶, they become the bottleneck of the research design

⁵One thing is certain: in terms of VR headset resolutions, there is a huge room for improvement (Orland, 2013). It is one thing to look into a FullHD/ 4K/ 8K display from the distance of one's chair/desk; having the display cover the whole field of view from an up-close distance is something else.

⁶A differentiation rate that is not an exponent of the framerate of the visualization device does introduce some marginal error as well.

and the related data output. E.g., eye-tracker resolution matters when targeting small/distant object; considering this, a question arises: what is considered an object small enough, or distant enough not to be implemented into a visual scene that tests for object (re)cognition?

4 Constructing the scene

Figure 1 shows an eye-tracked scene we created for the means of our research. The rest of this chapter will be devoted to the methodical approach we adopted to produce such a scene.



Fig. 1: An example scene with eye-tracked objects in it.

To actually have something worthy of measuring in a scene, the intended scene must be purposely designed in way of it being saturated in stimuli that are relevant to the theory we are following. In case of finding ways of differentiating from *spatial visualizers* and *object visualizers*, the scene must contain measurable elements that engage both of these kinds of cognitive styles, differently, and these differences can be recorded. On top of this, knowing that visual cognition is a continuous process, predicting possible recognition patterns these two types of visualizers may undertake has the potential of solidifying our method.

For a predominately object-oriented visualizer, outfitting the scene with objects rich enough in detail is crucial. When constructing and placing such objects into the scene, let them purposefully be:

- **Unique.** It is a common shortcoming of computer-generated 3D scenes that these scenes appear rather bland. This is so due to the creators of such scenes re-using the same 3D objects over and over again. Therefore, the object of focus shall be made of an unique 3D model; on the other hand, reasonable modesty in visual appropriateness should still be considered: when a scene is constructed within the constraints of a visual theme, any object within that scene should still respect that theme. There is

such a thing as an object breaking the overall impression of a theme; where the line of such over-intrusion lies, that is, however, unclear (albeit discoverable by pilot testing).

- **Breaking the pattern.** It is known that patterned visual stimuli, such as those found in nature, are not very demanding in terms of visual processing/recognition, or worthy of one's prolonged attention (Taylor et al., 2005). Therefore, an object of focus shall aim for coming up with a reasonably unique external shape (silhouette) and internal patterns. Let these be (somewhat) non-repetitive/non-conventional, too.
- **Detailed.** An object-oriented participant is inclined to gaze into the details of individual objects. To facilitate this, let the object have unique features of varying details. As it is with art, bland, visually simplistic areas shall complement detailed, somewhat unexpected parts. Such dynamics should be contained within an object.

Figure 2 shows some of the object-oriented elements placed into the scene. In red: a detailed flag. In blue: a unique object of an atypical chair. In yellow: a pattern-breaking set of bricks.

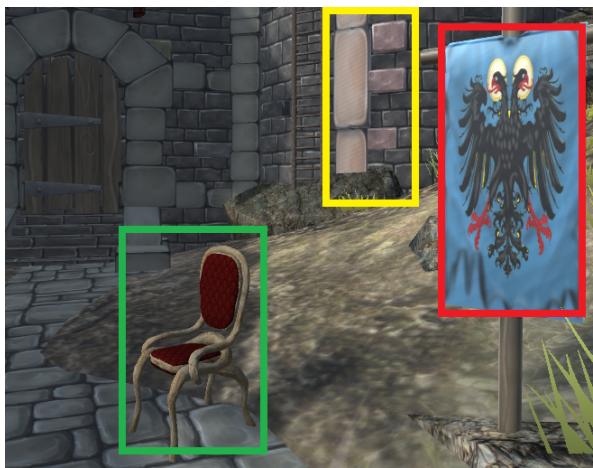


Fig. 2: Some of the object-oriented stimuli in the scene.

As for spatial visualizers, enriching the scene with relationships is the key. This globalistic/holistic individual does not spend much time on individual objects; instead, they search for rules that bound the elements of the scene into patterns. Therefore, let us consider these suggestions while making the scene stimulating for the spatial visualizer:

- **Lead lines.** In visual art, a piece of imagery conveys its message in a flow of lines (three-by-three division, two-by-two-division, one/two/three point perspective, horizon line, convergence of lines to

an “action center”, etc.⁷) It can be expected for the participant to follow these (guide)lines, to flow across the scene, or to proceed on to an area of focus (ideally, any scene should have more than one of these areas of focus, so as to prevent pro-object visualizer interpretation).

- **Scene boundaries.** These distinctive elements, if in place, should indicate where an intended scene ends. Let them take the form of massive and/or close-up monolithic objects situated on the fringes.
- **Similarities.** If there are multiple objects in the scene with distinctively similar features, these will form a logical group (despite potentially being intentionally displaced out of proximity). The spatial visualizer, being the one who scans across the whole scene faster than the object visualizer, is the prime candidate to notice the pattern of similarities, and to follow these groupings of objects.
- **Encompassion.** This can be achieved by placing a large object that spans across a significant area of the scene, followed by inserting supplementary objects into the body of the large one – all in a way that respects, and complements, the shape and the flow of the large object.

Figure 3 shows the spatial layout of the scene. In blue: prominent building being divided into three portions by lead lines, along with the horizon line. In violet: a surrounding valley that represents scene boundaries.

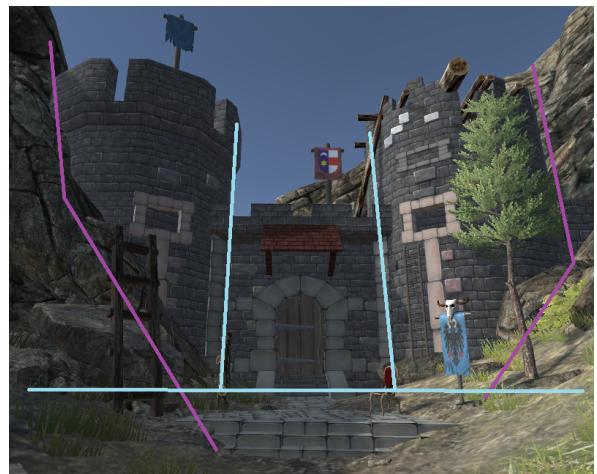


Fig. 3: The prominent spatial cues in the scene.

It is also worth considering that some differences between the two cognitive styles might not manifest outright; they may come to light only after pilot testing our scene (be it from self-reports of trial participants, or from analyzing their data).

⁷All these terms tie to art fundamentals theory. To explain these further, refer to any serious art textbook – such as this one: (Beloëil et al., 2013).

5 Optimizing the scene

The fact that we use object tracking means that if we, somewhat counter-intuitively, wish to gain extra data from an object in the scene, this object has to be broken into multiple sub-objects (while retaining its appearance of a whole object - the difference being merely functional, happening at a data-collection level).

To prevent inaccuracies and/or data noise, it is also recommended not to overlay eye-tracked 3D objects onto other 3D objects too much; especially with the added depth axis (present in the 3D image, obviously not present in eye-tracking data) it could be difficult to determine the object of focus, should edges of multiple objects frequently cross, especially along the depth axis. In fact, eliminating the depth axis to an extent is considered preferable; this can be achieved by placing the objects into the scene, followed by placing a wall, or some other monolithic barrier of constant depth, beyond these eye-tracked objects.

Here are some further considerations regarding constructing scenes for eye-tracking:

- **Long distance** object placement avoidance. Usually, some 30m may be considered maximum (depending on the overall size of the object).
- **Object size.** To ensure the objects in the scene are big enough, project the 3D scene onto a 2D plane; on this 2D plane, an object should take up at least 5% screen width and height to be considered accurately trackable and displayable within the bounds of current VR headset resolutions.
- **Proper scene lighting** should complement object recognizability.
- **Objects on the edge** of the scene will be implicitly harder to notice than objects in the middle; this goes double for objects that are away from the default field of view – if there are any (i.e. those objects only accessible by initial head movement up/down/to sides).
- **Static camera.** For increased control over the experimental design, do not allow the participants to move through the environment while viewing the eye-tracked scene. However, this solution is a trade-off that throws away movement data and possible interpretations related to it (i.e. participants actively positioning themselves in certain way to make sense of the scene).
- **The intended field of view of the scene**, and the potential need for camera rotation to view a widespread scene are to be considered. A scene may span across the default camera field of view (110 degrees, for most virtual reality headsets), or it can go further. A 360-degree scene is usually the

most demanding one for the participants to make sense of, remember and recall – because more object/relations can be stored in broader scenes.

Prior to an actual, in-depth data analysis (which is not the scope of this paper), data cleaning happens. In this case, adhering to the aforementioned suggestions does help with reducing the potential for data noise. When boundaries of few to none eye-tracked objects are neighboring, a number of quickly altering fixations (caused by inaccuracies in eye-tracking) are few as well. Furthermore, by adding some tolerance to object focus boundaries and defining minimum time intervals for eye-tracking fixations (as opposed to discardable eye saccades), data clarity can be achieved.

To analyze our data, we have no specialized software package to use (a situation unlike the conventional 2D eye-tracking). For most of our data analysis, we'll have to develop a solution of our own; to achieve this, a solution based on the Processing data visualization programming language may have to be created.

6 Supplementing our approach

While experiencing the 3D visualization, the participants can verbalize their thought processes; the verbalization data can be then coded into object/spatial meaning units. Another approach is verifying their remembrance post-test, by a means of presenting the participants with multiple-choice imagery of some cut-out portions of the scene, with one variant being correct (for an example of this, see Figure 4).

Asking an open question of “*What is that you noticed in the environment?*” does not prime the participants to answering under the influence of suggestive inquiry – their visual cognitive style preference may unveil naturally this way. As for object visualization disposition, the researcher can ask the participants to describe objects in the scene that they remember, to describe a specific object, or to probe for a specific feature or existence of an object (e.g. “*What did the coat of arms look like?*”, “*Was there a wheel in the scene?*”, etc.) The spatial-disposition questioning researcher asks for relations. They can pick an object in the scene and ask for associations to other objects (e.g. “*Was there a log right to the ladder?*”, “*What was the item highest up in the scene?*”, etc.). This sort of questioning can transcend even beyond the predetermined scenes, across the whole 3D environment.

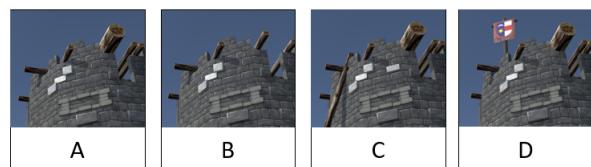


Fig. 4: Post-test questioning (the layout of a tower).

Another way of post-test questioning can be achieved by using maps. To test for spatial preference, the participants will be asked to draw/place objects into the map in accurate proportions/coordinates; as for object-oriented preferences, the participants may be presented with multiple variations of one item to place on the map (one correct variation, the rest of them incorrect).

To make relevant comparison of our findings, our participants were asked to complete the *Object-Spatial Imagery questionnaire* on the side; the intention was to compare the results of this questionnaire side-to-side with our our data.

7 Conclusion

Utilizing the existing technology and some scripting of our own, we have developed a platform that allows for collecting behavioral data in virtual environments; within the scope of such environments, we have proposed a way of creating eye-trackable 3D scenes that can be used in visual cognitive styles testing. The proposed process of creation was intended to respect the visual cognitive styles theory. Also, potential technological shortcomings were covered – so that the processes of data collection and data cleaning are not negatively affected.

We are yet to analyze data and produce results based on an eye-tracking task we have implemented. The approach we have proposed in this paper can, however, serve as a research design baseline for future studies in visual cognitive styles (or in similar subjects).

Acknowledgment

This submission was made possible due to the following project: Vliv kartografické vizualizace na úspěšnost řešení praktických a výukových prostorových úloh (MUNI/M/0846/2015).

References

- Ausburn, L. J., & Ausburn, F. B. (1978). Cognitive styles: Some information and implications for instructional design. *Educational Communications & Technology Journal*, 26, pp. 337-354.
- Basu, A., Ball, C., Manning, B., Johnsen, K. (2016). Effects of User Physical Fitness on Performance in Virtual Reality. *2016 IEEE Symposium on 3D User Interfaces*, pp. 233-234.
- Beloceil, G., Riabovitchev, A., Castro, R. F. (2013). Art Fundamentals: Color, Light, Composition, Anatomy, Perspective, and Depth. 3DTotal Publishing.

Blajenkova, O., Kozhevnikov, M., Motes, M. A. (2006). Object-Spatial Imagery: A New Self-Report Imagery Questionnaire. *Applied Cognitive Psychology*, 20, pp. 239–263.

Dolezalova, J., Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. *Journal of Eye Movement Research*, 9 (4), pp. 1-13.

EyeSee3D website. Retrieved from <http://eyesee3d.eyemovementresearch.com/>

Kozhevnikov, M. (2007). Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style. *Psychological Bulletin*, 133 (3), pp. 464–481.

Kozhevnikov, M., Kosslyn, S., Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, 33 (4), pp. 710-726.

Howarth, P. A., Costello, P. J. (1997). The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18 (2), pp. 107-116.

Lamkin, P. (2017, March 14). Best VR headsets 2017: HTC Vive, Oculus, PlayStation VR compared. Retrieved April 07, 2017, from <https://www.wearable.com/vr/best-vr-headsets-2017>

Object-Spatial-Verbal Cognitive Style Model. Retrieved from <http://www.nmr.mgh.harvard.edu/mkozhevnlab/?tag=object-spatial-verbal>

Orland, K. (2013, September 11). Virtual Perfection: Why 8K resolution per eye isn't enough for perfect VR. Retrieved April 07, 2017, from <https://arstechnica.com/gaming/2013/09/virtual-perfection-why-8k-resolution-per-eye-isnt-enough-for-perfect-vr/>

Stanney, K. M., Kennedy, R. S. (2008). Simulation Sickness. *Human Factors in Simulation and Training*, pp. 117-127.

Taylor, R. P., Spehar, B., Wise, J. A. et al. (2005). Perceptual and Physiological Responses to the Visual Complexity of Fractal Patterns. *Nonlinear Dynamics, Psychology, and Life Sciences*, 9 (1), pp. 89-114.

Nový pohled na výpočty a umělá inteligence

Jiří Wiedermann

Ústav informatiky AV ČR

Pod Vodárenskou věží 2, 182 07 Praha, Česká republika

a

Český institut informatiky, robotiky a kybernetiky

ČVUT Praha, Česká republika

Email: jiri.wiedermann@cs.cas.cz

Abstrakt

Klasický pohled na výpočty je chápán jako procesy generované počítačem, tj. soustředí se na to, JAK jsou výpočty realizované. Nový pohled se zajímá o to, CO výpočty dělají, co je jejich smyslem — a to je generování znalostí. Výpočty tudíž chápeme jako procesy, které generují znalosti nad danou znalostní doménou v rámci příslušné znalostní teorie. Inteligence je pak schopnost získávat informace a transformovat je na znalosti, které jsou dále využívány pro řešení problémů. Ukážeme, že toto pojednání výpočtů umožní přirozeným způsobem definovat některé vyšší kognitivní funkce jako je zdůvodňování svých akcí, sebe-uvědomění, introspekce, porozumění znalostem, svobodná vůle, kreativita a také porozumění algoritmickým mechanismům, které stojí za rozvojem inteligence. Tento pohled je přínosný tím, že na elementárním abstraktním modelu kognitivního systému, který není zatížen žádnými technickými detailemi, ukazuje, že všechny výše zmíněné kognitivní funkce souvisejí se specifickými znalostmi, které jsou generované v rámci téhož modelu.

1 Úvod

Výpočty jsou tradičně chápány jako procesy, které probíhají v počítačích. Pokud bychom chtěli tuto definici výpočtu zpřesnit, musíme zpřesnit pojem počítače. Za tím účelem se v teorii obyčejně uvažuje Turingův stroj, o kterém je známo, že tento model v principu zachycuje výpočetní schopnosti velké třídy současných digitálních počítačů. Nicméně, v praxi, a zejména v umělé inteligenci nebo v biologii, často považujeme za výpočty i procesy, které probíhají nikoliv v lidmi uměle zhotovenými zařízeními (tj. v počítačích), ale také např. v mozku lidí či zvířat, v rostlinách anebo buňkách. Ve fyzice se vyskytují úvahy, že dokonce na celý vesmír lze nahlížet jako na obrovský počítač, který se řídí „instrukcemi“, které představují fyzikální zákony. Takže pokud se chceme dozvědět něco o výpočtech, musíme zkoumat schopnosti a meze těchto zařízení. To znamená, že se soustředíme na problém, JAK jsou výpočty realizované. Takový pohled je sice technicky zajímavý, avšak vede na (klasickou) teorii výpočtů, která je stro-

iově závislá a nemá potenciál zachytit, jaký vlastně je smysl, cíl počítání — co vlastně výpočty „dělají“ či mohou „dělat“, ať už pro nás, pro lidi, anebo pro to zařízení, které výpočet realizuje.

Pokud se tedy ptáme, co výpočty vlastně dělají, tak jedinou smysluplnou odpověď je, že produkují znalosti. Toto je výchozí bod tzv. *znalostní teorie výpočtů*, která vychází z prací autorů van Leeuwena a Wiedermannova (2013, 2014, 2015a, 2015b, 2017). Dle této teorie jsou výpočty chápány jako procesy, které generují znalosti nad danou znalostní doménou v rámci příslušné znalostní teorie. V našem příspěvku ukážeme, že tento pohled má velký potenciál zejména pro umělou inteligenci, protože umožňuje přirozeným způsobem definovat (a chápout) netriviální kognitivní funkce jako je zdůvodňování svých akcí, sebe-uvědomění, introspekce, porozumění znalostem, svobodné vůli a kreativitě a přináší vhled do algoritmických mechanismů, které stojí za rozvojem inteligence. Toto lze považovat za zásadní přínos k problematice výpočetních kognitivních systémů, protože jiné známé přístupy k popisu těchto funkcí používají daleko složitější modely kognitivních systémů než je náš znalostní model. Tyto jsou obvykle tak složité, že výsledný popis či definice kognitivních funkcí opět mají charakter, který spíše záleží na vlastnostech daného modelu, na jeho architektuře nežli na obecných vlastnostech uvažovaných funkcí. Podrobný přehled realizovaných kognitivních architektur lze nalézt v práci a v on-line katalogu Samsonoviche (2010).

Struktura příspěvku je následující. V části 2 uvedeme základní ideje znalostního přístupu k výpočtům. V části 3 vysvětlíme, jak nový pohled na výpočty přirozeným způsobem vede k definici základních vyšších kognitivních funkcí, jakými jsou důvodění, sebeuvědomění, introspekce, porozumění, svobodná vůle, kreativita a sebe-zdokonalování znalostních teorií. Poslední část 4 shrnuje základní poznatky našeho příspěvku.

2 Výpočet jako generování znalostí

V souladu se znalostní teorií výpočtu budeme nahlížet na znalost jako na výsledek nějakého výpočetního procesu, který pracuje nad jistou znalostní doménou tak, že kombinuje její prvky — *elementární znalosti* — do odvozených, často složitějších konstrukcí, které tvoří novou znalost, opět nad danou doménou. Pro kombinaci těchto prvků používá výpočet množinu (odvozovacích) pravidel, která může být předem daná, anebo se může tvořit pomocí učení během velkého počtu různých výpočtů nad danou doménou. Inteligentní systém tímto způsobem pracuje s více či méně formální teorií, která zachycuje vlastnosti dané znalostní domény a způsoby odvozování nových znalostí, stále v rámci dané domény.

Jak jsme již zmínili, přitom nás nebude v první řadě zajímat, JAK daný výpočet probíhá, ale CO výpočet počítá — jaká znalost je generována v průběhu výpočtu. Pod tímto zorným úhlem se stává schopnost generovat znalosti poznávacím kritériem těch procesů, které budeme nazývat jako výpočetní procesy neboli výpočty. Inteligentní systémy jsou tudíž speciálními případy výpočetních systémů, u kterých je schopnost generovat znalost maximalizována v tom smyslu, že takové systémy jsou schopny generovat znalost nad libovolnými znalostními doménami modelujícími velké části reálného světa anebo různých věd. To je v protikladu s praxí současných AI systémů, které jsou zpravidla specializovány na specifické, většinou značně omezené znalostní domény. Procesy, které negenerují znalost, nebudeme považovat za výpočty. Zde narázíme na problém — jak poznáme, co je znalost? To je fundamentální filozofická otázka, na kterou filozofové dosud nenalezli odpověď. Jedno je jisté — co se jednomu jeví jako znalost, pro druhého to může být buď samozřejmost, anebo to nemusí vůbec považovat za znalost. Pojem znalosti není tedy absolutní, ale je *závislý na pozorovateli* (*observer dependent*). Závisí tedy na tom, co již pozorovatel zná. Ve znalostním přístupu se to řeší tak, že znalost se definuje v rámci nějaké *zalostní domény*, nad kterou výpočet operuje. Všechny znalosti o nějaké podmnožině znalostní domény jsou zachyceny pomocí *zalostní teorie*, která může být více či méně formální, anebo zcela neformální. V rámci této teorie popisují *axiomy* elementární znalosti, které odpovídají (reprezentacím) objektů ve znalostní doméně a jejím vlastnostem. Způsoby, jak lze z takových elementárních znalostí konstruovat nové, odvozené znalosti jsou popsány pomocí *odvozovacích pravidel*. Výpočetní procesy jsou svázány s odpovídající znalostní doménou prostřednictvím znalostní teorie, se kterou výpočty přímo či nepřímo pracují, pomocí následující podmínky: *cokoliv lze odvodit v rámci dané teorie musí být podporováno příslušným výpočetním procesem*. Pokud je tomu tak, pak to, jaká znalost může anebo nemůže být generována nad danou znalostní doménou, a „kvalita“ takto generované znalosti (tj. např. její shoda s pozorováním) závisí výlučně na

vlastnostech odpovídající znalostní teorie.

Všimněme si, že znalostní přístup k výpočtu je *strojově nezávislý*, protože platí pro jakýkoliv proces realizující odvozování v rámci dané znalostní teorie. Taktéž je *algoritmicky nezávislý*, protože se nezajímáme o to, jakými postupy je výpočetní proces realizován. V neposlední řadě je také *nezávislý na reprezentaci*, protože v našem přístupu nepředpokládáme žádnou speciální reprezentaci znalostí.

Díky své obecnosti znalostní přístup lze uplatnit nejen v dobře formalizovatelných, tzv. *exaktních znalostních doménách*, ale i ve znalostních doménách a pro odvozovací pravidla, které se vzpírají jakékoliv formalizaci. Takovým doménám budeme říkat *popisné znalostní domény*. Typickým případem popisné domény s neformálními odvozovacími pravidly je reálný svět. Jeho objekty, jevy, akce a vztahy mezi jimi jsou popsány pomocí přirozeného jazyka. Znalosti o takové doméně jsou zachyceny ve větách přirozeného jazyka. Odvozovací pravidla jsou v tomto případě tzv. *pravidla racionálního uvažování a chování*. Tato pravidla vycházejí z faktů a argumentací, která lze zachytit v přirozeném jazyce. V typickém případě mají popisné domény rozsáhlé znalostní báze (jako např. obsah internetu) a relativně krátké odvozovací řetězce.

Význačným příkladem inteligentního systému vykazujícím lidskou inteligenci je mozek společně s přirozeným jazykem. Mozek umožňuje odvozovací procesy v neformální teorii, kterou lze popsat v přirozeném jazyce. Samozřejmě, že v principu můžeme namísto mozku uvažovat jakýkoliv počítač s podobnými vlastnostmi, i takový, o kterém doposud nevíme, že existuje; výsledkem bude systém umělé inteligence na lidské úrovni. Skutečnost, že znalostní přístup k výpočtu umožňuje pracovat i s takovými nedokonale definovanými pojmy je předností našeho modelování. To nám umožní dosáhnout nové porozumění problematice generování znalostí, které zatím nebylo dosaženo jiným způsobem. Pro formalizaci tohoto přístupu viz práci van Leeuwena a Wiedermannu (2017). Přehled dosavadních výsledků z oblasti kognitivních výpočtů lze nalézt v práci Wiedermannu a van Leeuwena (2015a).

3 Kognitivní funkce jako nadstavba nad mechanismem generováním znalostí

Vraťme se teď zpět k podmínce, která svazovala výpočetní proces se znalostmi, které lze nad danou znalostní doménou odvodit. Zde jsme požadovali, aby jakákoli znalost, kterou lze odvodit (formou důkazu v dané znalostní teorii nad danou znalostní doménou) byla prokazatelně (tj. opět: musí existovat důkaz) podporovaná příslušným výpočetním procesem.

Pokud umělý kognitivní systém pracuje tímto způsobem, může při vhodné organizaci své činnosti realizovat některé netriviální vyšší kognitivní funkce, jejichž definice není na první pohled zřejmá, a tím spíše

jejich realizace.

Důvodění (accountability) Důvodění znamená schopnost zdůvodnění svých rozhodnutí, tj. kognitivní agent může na vyžádání podat zdůvodnění svých akcí, vysvětlení „jak na to přišel“, viz Kroll a spol. (2016). Za tím účelem stačí, aby agent společně s výslednou znalostí generoval i důkaz, který použil při jejím odvozování. Tento důkaz již samozřejmě může být prezentován ve formalismu, kterému uživatel rozumí. Takový důkaz umožní uživateli „kontrolu správnosti“ výsledku v rámci odpovídající znalostní teorie v porovnání s tím, co uživatel očekával na základě svého zadání.

Sebeuvědomění (awareness) Díky schopnosti důvodění znalostní systém „ví“ (má informaci), jaký problém řeší a je schopen podat vysvětlení, jak ho řeší. Znalost o sobě může být součástí znalostní teorie, která řídí činnost systému.

Introspekce Na základě skutečnosti, že si kognitivní systém pamatuje své předchozí řešené úkoly a způsoby jejich vyřešení, může se k ním vracet, znova je podrobit zkoumání a využít je při řešení nových problémů pomocí analogie anebo vylepšit jejich původní řešení s ohledem na nové poznatky, které systém mezičítím mohl získat.

Znalostní porozumění Schopnost důvodění, sebeuvědomění a introspekce dohromady představují schopnost porozumění znalostní doméně, s kterou systém pracuje. Systém je schopen vysvětlit význam termínů, které používá a na základě své předchozí zkušenosti (které jsou zapamatovány v příslušné znalostní bázi) kreativně je aplikovat v novém kontextu. Pro plné porozumění reálnému světu je potřebné uvažovat vtělené systémy.

Svobodná vůle Budeme říkat, že kognitivní systém A má svobodnou vůli právě tehdy, když neexistuje kognitivní systém B, který výlučně na základě pozorování chování systému A v různých situacích dovede vždy předpovědět chování systému A v dané situaci. Tato naše definice se liší od standardních definicí (kterých je nepřeberné množství — viz např. odpovídající heslo na Wikipedii) a které vidí svobodnou vůli jakoby z vnitřního pohledu systému, tj. subjektivně. Např. „svobodná vůle je schopnost zvolit si různé možnosti chování v dané situaci,“ anebo „schopnost chovat se ve výsledku jinak než na základě minulých událostí“. O tom, že si kognitivní systém zvolil z různých možností chování, anebo že se chystá zahovat jinak, než předtím, má informaci pouze systém samotný. A protože v obecném případě systému „nevidíme do hlavy“, nejsme schopni rozhodnout, jestli má svobodnou vůli. Je však zřejmé, že oba právě zmíněné případy (a další) zachycuje naše definice tak, že vnější pozorovatel není schopen predikovat budoucí akce systému. To znamená, že naše definice činí pojem svobodné vůle závislým na pozorovateli. Má však

výhodu v tom, že činí problém, jestli má systém svobodnou vůli, rozhodnutelným (samozřejmě vzhledem k pozorovateli).

Kreativita Kreativita je projevem kreativního procesu, což je každý proces, který generuje znalost řešící problém, jenž je pro kognitivní systém nový. Je to protiklad rutinního procesu řešícího známý problém pomocí již známých postupů. Obecně, řešení problému vyžaduje najít znalost, která splňuje předem danou množinou podmínek. Jinými slovy, hledáme explicitní znalost, která je implicitně zadaná pomocí vlastnosti, které musí tato znalost splňovat. Pokud tento úkol modelujeme pomocí znalostního přístupu k počítání, tak výchozím postupem pro hledání řešení zadaného úkolu je použití „hrubé síly“ — systematického generování všech znalostí, které lze v systému (tj. v příslušné znalostní teorii) odvodit a zkoušení, jestli daná znalost nesplňuje zadané podmínky. To vypadá jako hrubě neefektivní až naivní přístup, ale zdá se, že tento přístup je v pozadí jakéhokoliv kreativního procesu. Jedná se tedy o speciální případ objevování znalostí (*knowledge discovery*). Naštěstí, opakováním používáním takového zpočátku neefektivního, leč univerzálního postupu jej lze kultivovat. Na objevování znalostí hledíme jako na jeden interaktivní, evoluční a potenciálně nekonečný sebezdkonalující se a učící se proces, jehož cílem je zlepšovat své vlastní kreativní schopnosti. Kultivaci procesu objevování znalostí lze popsat Wiedermannem a van Leeuwensem (2015b) dosahovat zejména

- iterativním zjemňováním a/nebo rozšiřováním vyhledávacích kritérií na základě předchozích zkušeností a/nebo málo úspěšných pokusů;
- automatickou extrakcí a modifikací uživatelských kritérií, které zajišťují výběr možností na základě odpozorovaných preferencí uživatele a jeho emocí a prožitků (podobně, jako to dělá Google+). Toto se děje při každé příležitosti (tj. nejen pro řešení každého konkrétního problému zvlášť) a získané poznatky jsou využívány pro usměrňování jakéhokoliv objevného procesu a pro uspořádávání objevovaných výsledků.
- využíváním předchozích zkušeností, které mohou být odvozeny z epizodické paměti;
- řízenou interakcí s okolím zaměřenou na získání dalších znalostí, které mohou být nápomocny při řešení konkrétního problému (podobně, jako když na Internetu vyhledáváme dodatečné informace).

Použitím takových kultivačních procedur se mění (doplňuje) jak systémová báze znalostí, tak i mechanismy práce s ní.

Sebe-zdokonalování znalostních teorií Samotnou schopnost odvozovat další znalosti v rámci dané znalostní teorie však nelze považovat za hlavní znak inteligentních systémů. Tím je až jejich schopnost vylepšovat svoji znalostní teorii, pomocí a v rámci které

tyto systémy generují své znalosti. Inteligence takových systémů prokazatelně roste, protože získávají kvalitativně nové znalosti o znalostní doméně, nad kterou pracují.

Při použití kultivačních procedur popsaných v předchozím případě se může stát, že systém získá novou znalost, které je ve sporu se znalostmi, které již systém má. Bud' si takovou znalost odvodí systém sám, anebo ji získá „zvnějšku“ (třeba z Internetu) anebo vlastním pozorováním a odhalením nesouladu pozorování s vlastní teorií. Pro takové případy musí mít vyspělý systém mechanismy, které odhalí logickou inkonsistenci své znalostní teorie. Takovou vadu lze odstranit jedině změnou příslušné teorie. Systémy, které jsou schopny sebe-zdokonalovat svoji znalostní teorii, a jsou navrženy tak, aby systematicky vyhledávaly rozporu své teorie s fakty, zřejmě mohou automaticky, pomocí svých vlastních mechanismů, zvyšovat svoji vlastní inteligenci, při jakékoli rozumné definici pojmu intelligence. Takové sebe-zdokonalování může pokračovat do té doby, pokud existují rozporuplná fakta a systém je objeví, a také pokud ve znalostní doméně existují nové, doposud neprozkoumané objekty a jevy (Bostrom 2014). Takové systémy pak mohou, ale spojen teoreticky a v některých směrech, překonat inteligenci lidskou (Wiedermann a van Leeuwen 2017). Pojem sebe-zdokonalovacích znalostních teorií překonává svým dopadem dosavadní obecné představy o tzv. sebe-zdokonalovacím software (viz např. Bostrom 2014), protože identifikuje jako nutnou podmínu pro nárůst intelligence kognitivních systémů znalostní data, jejich kvalitu i kvantitu (a tedy — kvalitu příslušných znalostních teorií) a nikoliv efektivitu odpovídajících odvozovacích mechanizmů.

4 Závěr

V příspěvku jsme ukázali nový pohled na výpočty, který je chápě jako procesy generující znalosti a využívá tento pohled k definici a pochopení netriviálních vyšších kognitivních funkcí. Takovými funkciemi jsou důvodění, sebeuvědomění, introspekce, porozumění, svobodná vůle, kreativita a sebe-zdokonalování znalostních teorií. Tyto funkce nelze elegantně popsat pomocí klasického pohledu na výpočty, který považuje za výpočty jakékoli, i nesmyslné procesy, generované různými modely počítačů. Takový pohled je nutně strojově orientovaný a tudíž neposkytuje dostatečně abstraktní a obecný rámec pro definici zmiňovaných kognitivních funkcí a jejich pochopení. Naopak, nás pohled přes teorii kognitivních výpočtů ukazuje pomocí elementárního abstraktního modelu kognitivního systému, který není zatížen žádnými technickými detailemi, že všechny výše zmíněné kognitivní funkce souvisejí se specifickými znalostmi, které jsou generované v rámci téhož modelu.

Poděkování

Tento příspěvek vznikl za částečné podpory GA ČR v rámci grantové úlohy 15-04960S, institucionálního plánu ÚI AV ČR RVO 67985807 a programu Strategie AV21.

References

- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press
- Kroll, J.A. a spol. (2016). Accountable Algorithms (March 2, 2016). University of Pennsylvania Law Review, Vol. 165, 2017 Forthcoming; Fordham Law Legal Studies Research Paper No. 2765268. Available at SSRN: <https://ssrn.com/abstract=2765268>
- Samsonovich, A.V. (2010). Toward a Unified Catalog of Implemented Cognitive Architectures. BICA 221 (2010): 195-244, on-line katalog <http://bicasociety.org/cogarch/>.
- van Leeuwen, J. a Wiedermann, J. (2017). Knowledge, representation and the dynamics of computation. In: G. Dodig-Crnkovic, R. Giovagnoli (Eds): *Representation and Reality: Humans, Animals and Machines*. Berlin: Springer
- Wiedermann, J. a van Leeuwen, J. (2013). Rethinking computation. In: Proc. 6th AISB Symp. on Computing and Philosophy: The Scandal of Computation - What is Computation?, AISB Convention 2013 (Exeter, UK), AISB, pp. 6-10
- Wiedermann, J. a van Leeuwen, J. (2014). Computation as knowledge generation, with application to the observer-relativity problem. In: Proc. 7th AISB Symposium on Computing and Philosophy: Is Computation Observer-Relative?, AISB Convention 2014 (Goldsmiths, University of London), AISB, 2014
- Wiedermann, J. a van Leeuwen, J. (2015a). What is Computation: An Epistemic Approach. (Invited talk). In: Italiano, G. et al., (eds.). SOFSEM 2015: Theory and Practice of Computer Science. LNCS 8939, Berlin: Springer, pp. 1-13
- Wiedermann, J. a van Leeuwen, J. (2015b). Towards a Computational Theory of Epistemic Creativity. In: Proc. 41st Annual Convention of AISB 2015. London, pp. 235-242
- Wiedermann, J. a van Leeuwen, J. (2017). Understanding and Controlling Artificial general Intelligence Systems. In: Proc. 10th AISB Symposium on Computing and Philosophy: Language, Cognition and Philosophy, AISB Convention 2017, (University of Bath, UK), AISB, s. 356-363

Register autorov

A

- A. Franz, Elizabeth 14
Andrejková, Gabriela 130, 178

B

- B. Klein, Frederico 155
Beňuš, Štefan 8
Beňušková, Ľubica 14
Budinská, Ivana 24

C

- Cangelosi, Angelo 155

Č

- Čavojová, Vladimíra 30

D

- Démuth, Andrej 37
Dudeková, Katarína 46, 93

F

- Farkaš, Igor 104
Flegr, Jaroslav 169

G

- Gáliková, Silvia 52
Gergeľ, Peter 60

H

- Hlaváčková-Schindler, Kateřina 99
Hoffmann, Matěj 155
Hrebeňárová, Bernadeta 136

J

- J. Shaw, Daniel 141
Jáni, Martin 71
Jedlička, Peter 76
Juřík, Vojtěch 81, 184
Jurkovič, Marek 30

K

- Kamhal, Dezider 87

Konečný, Matúš 46, 93

Kopčo, Norbert 116, 136, 178

Kostovičová, Lenka 46, 93

Kováč, Anton 99

Kovács, Peter 104

Kozák, Pavol 130

Kuzmíková, Jana 110

L

Lúčny, Andrej 123

Lokša, Peter 116

M

M. Havel, Ivan 64

McCarthy, Paul 14

Mitášová, Monika 64

S

Salem, Asmaa 130

Straka, Zdeněk 155

Svoboda, Aleš 163

Š

Šašinka, Čeněk 81, 184

Šebeňa, René 136

Špliláková, Beáta 141

Šrol, Jakub 149

Štěpánová, Karla 155

T

Toman, Jan 169

Tomoriová, Beáta 178

U

Ugwitz, Pavel 81, 184

V

Vavrečka, Michal 155

W

Wiedermann, Jiří 190