



Kognice a umělý život **XVI**

Sborník z 16.ročníku konference Kognice a Umělý život (KUZ XVI)

Telč, Česká republika, 1.-3.6. 2016

Editoři: Michal Vavrečka, Ondřej Bečev, Matěj Hoffmann, Karla Štěpánová

June 7, 2016

- **Připraveno:** ČVUT v Praze
- **Kontakt:** Michal Vavrečka, Karlovo náměstí 13, CIIRC CVUT v Praze
- **ISBN:** 978-80-01-05915-9
- Sborník je také dostupný online na: <http://www.kognice.eu/>

Obsah

- 1 SOM model neurónovej siete v detekcii anomálií v textoch
Abdulwahed Almarimi, Gabriela Andrejková, Peter Sedmák
- 7 The temporal profile of contextual plasticity in sound localization
Gabriela Andrejková, Norbert Kopčo
- 13 Konceptuální analýza vědomí
Bc. Jiřina Babická
- 19 Vedecké intuície a vedecká gramotnosť u slovenských stredoškóľákov
Vladimíra Čavojová
- 23 Čo z čoho a kto z koho? Efekt číselného vyjadrenia a odbornosť pri posudzovaní rizika v oblastiach zdravia a financií
Katarína Dudeková, Lenka Kostovičová
- 29 On abstraction: psychological, neural, and computational perspectives
Igor Farkaš
- 37 Modelling subjective experience
Silvia Gáliková
- 49 Omylnosť metakognície o dôvodoch jednání a její trénink
Ondřej Havlíček, Marcel Brass, Axel Cleeremans, Agnieszka Wykowska
- 55 The encoding of proprioceptive inputs in the brain: knowns and unknowns from a robotic perspective
Matěj Hoffmann, Nad'a Bednářová
- 67 Funkčné zmeny mozgu počas spracovávania sociálnych informácií u schizofrenie: multimodálna meta-analýza neurozobrazovacích štúdií
Martin Jáni, Tomáš Kašpárek
- 71 Vyššie emócie: neurovedecký a psychologický pohľad
Dalibor Jurášek, Peter Marman
- 77 Dôvera a nedôvera vo vedecké a nevedecké informácie
Marek Jurkovič
- 83 Když je více méně: Význam lidské chyby při interakci s 3D virtuálním geografickým prostředím
V. Juřík, Č. Šašínska
- 89 Finančníci - pragmatici? Konverzačný kontext a doménová špecifickosť v posudzovaní rizika a rozhodovaní
Lenka Kostovičová, Katarína Dudeková
- 95 Who's moving it? Sense of agency in schizophrenia
Eva Kozáková, Eduard Bakštein, Filip Španiel, Tereza Nekovářová, Jan Rydlo
- 101 Porozumenie literatúre
Jana Kuzmíková
- 105 Modelling of the Reference Frame of the Ventriloquism Aftereffect
Ing. Peter Lokša, doc. Ing. Norbert Kopčo PhD.
- 111 Riadenie simulovaného robota iCubSim pomocou architektúry Agent-Space
Andrej Lúčny, Matúš Kopernický
- 119 Implementace a synchronizace Hayekova stroje pro řízení robota Khepera IV
Lukáš Mamula
- 125 Cognitive stress during driving task
Ján Marek, Marek Otáhal, Olga Štěpánková
- 131 Aproximální numerický systém u dětí
Michala Plassová, Michal Vavrečka, Michael Tesař

- 139 Špecifiká kognitívneho výkonu pacientov po ložiskovej ischémii mozgu
Ján Rybár, Barbora Cimrová, Igor Farkaš, Monika Varga Doležalová, Roman Rosipal
- 147 „Řízená náhoda“ v metakreativitě
Aleš Svoboda
- 153 EEG detekce emoční valence: Frontální alfa asymetrie jako klasifikátor valence emocí
Bc. Michael Tesař, Mgr. Michal Vavrečka, PhD.
- 157 Na cestě k uměle inteligentním systémům: Vyhodnocování inteligence umělých systémů
Ondřej Vadinský
- 165 Znalostní meze (super)inteligentních systémů
Jiří Wiedermann

Self Organizing Maps in Text Anomalies Detections

Abdulwahed Almarimi, Gabriela Andrejková, Peter Sedmák

Institute of Computer Science, Faculty of Science

P. J. Šafárik University in Košice

Email: abdoalmarimi@gmail.com, gabriela.andrejkova@upjs.sk, peter.sedmak@student.upjs.sk

Abstract

Each written text in some language has some author or more authors (authors have their individual sublanguage). An analysis of some text if authors are not known could be done using methods of data analysis, data mining, and using structural analysis. In the paper, it is described a system of modified *Self-Organizing Maps* working on sequences built from a text. The system is trained to input sequences and after the training it determines text parts with anomalies using a cumulative error and a special layer of neurons.

1 Introduction

According to Chomsky (2005), there exist three factors that contribute to the growth of the individual language (assuming that the faculty of language has the general properties of other biological systems): (1) genetic endowment, (2) experience, which leads to variations, and (3) principles not specific to the faculty of language. An analysis of some text if authors are not known could be done according the third principle using methods of data analysis, data mining, and using structural analysis (a grammar of the used language).

A given text written in some genre, in some language and grammar presents a sequence of letters, sequence of words, sentences, sections and we find some anomalies in this text (written in a time). For example, anomalies should show that some parts of the text were written (modified) by another author or that somebody manipulated with the text. The problem belongs to problems working with text and studying authorship attribution and plagiarism, but in both problems there exist some groups of comparable authors and comparable texts. It means the results of analysis can be compared according to texts or authors. In our problem any author is known and we analyze each text as one extra text.

In our analysis, we will work with Self-Organizing Maps (SOM). Very good description of SOM extensions for temporal structures is in Hammer et al. (2005), some of the extensions are useable for sequences. In Durgin a Zhang (2005) is built a prototype anomaly detection tool that shows an integration of SOM model and ART model in an operational intrusion detection.

The plagiarism and authorship attribution problems were studied by Hassan a Chaurasia (2012), Stamatatos (2006b), Stamatatos (2010) using n -grams. Self Organizing Maps (SOM) were used by Neme et al. (2015) to stylistic analysis and authorship attribution.

The time series were studied by Barreto a Aguayo (2009) using SOM model. It is our new idea to apply the method to sequences built from texts.

In this paper, we applied a new our system for anomalies detection in Arabic and English texts based on SOM model neural networks.

The paper has the following structure: In the second section is given very short description of SOM model neural network from sequences learning point of view. In the third section we give information on some analysed texts, the fourth section contains a description of our developed system. The following section contains results and an evaluation of the system. In the conclusion, we summarise our results and give some plan for the following work in the area.

2 Self Organizing Maps

The Self Organizing Map belongs to the class of unsupervised and competitive learning algorithms. This type of neural network is used to map the n -dimensional space to the less dimensional space, usually to two dimension space. The neurons are arranged usually to the two dimensional lattice, frequently called a map. This mapping is topology safed and each neuron has its own n -dimensional weights vector to an input. If input is represented by some sequence (for example, time series), when the order of values is important, then it is necessary to follow the order and do not change it.

The steps of the algorithm:

1. **Initialization.** The weight vectors of each node (neuron) in the lattice are initialized to a small random value from the interval $\langle 0, 1 \rangle$. The weight vectors are of the same dimensions as the input vectors.
2. **Winner identification for an input vector.** Calculate the distance of the input vector to the weight vector of each node. The node with the shortest distance is the winner. If there are more than

one node with the same distance, then the winning node is chosen randomly among the nodes with the shortest distance. The winning node is called the Best Matching Unit (BMU). Let i^* be index of the winning node.

3. **Neighbours calculation.** For this is used the equation:

$$h(i^*, i, t) = \exp\left(-\frac{\|r_i(t) - r_{i^*}(t)\|}{\sigma^2(t)}\right), \quad (1)$$

where $\sigma(t)$ means the radius of the neighbourhood function, t is an iteration step, $r_i(t)$ and $r_{i^*}(t)$ are the coordinates of units i and i^* in the output array.

4. **Weights adaptation.** Only of the weights of the nodes within the neighbourhood radius will be adapted. the equation for that is :

$$\vec{w}^{new} = \vec{w}^{old} + \eta * h(i^*, i, t) * (\vec{x} - \vec{w}^{old}), \quad (2)$$

where \vec{w}^{new} is the vector of the new weights, \vec{w}^{old} are old weights, $\eta \in (0, 1)$ is the learning rate, \vec{x} is the actual input vector.

After the algorithm make changes in the weights, it presents next random input vector from the remaining input vectors to input and continue in the step 2 and so on until there is none input vector left.

3 Data Analysis, Learning Sequencies of Texts

In the text analysis we use English recommended texts from benchmark CorpusEnglish (2011) and Arabic texts from CorpusArabic (2011). In Table I we describe some information about 2 Arabic and 2 English texts, the texts A14 and E777 were constructed as a combination of two different texts (important for an illustration of our analysis). The letter position of the connection in both texts is shown in the results of the analysis.

We will use the texts as sequencies of letters and words. The time sequencies learning using SOM models were described in Barreto a Aguayo (2009).

Notation:

- Γ - a finite alphabet of letters; $|\Gamma|$ is the number of letters in Γ ; in our texts, Γ_A will be Arabic and Γ_E English alphabet;
- V - a finite vocabulary of words in the alphabet Γ presented in the alphabetic order; $|V|$ - the numbers of different words in the vocabulary V ;

3.1 Sequencies of letters and $n - gram_l$ s of letters

The text is a sequence of letters in the alphabet Γ . The numbers of letters occurencies are different in texts. We

Tab. 1: Statistics of 2 English and 2 Arabic texts, the number of words by length for 1 – 10 and maximal frequencies of 3 and 4 – *grams*.

| | Name of Texts | | | |
|-------------------|-----------------------|----------------------|-----------------------|----------------------|
| | A4 | A14 | E14 | E777 |
| Total-words | 31656 | 2168 | 40721 | 1655 |
| Total-letters | 135573 | 11409 | 216677 | 10407 |
| # diff. words | 10098 | 1108 | 7492 | 598 |
| # words by length | | | | |
| 1 | 81 | 28 | 1466 | 42 |
| 2 | 4816 | 312 | 5636 | 252* |
| 3 | 7795 24.28% | 529 24.40% | 9041 24.91% | 296 17.88% |
| 4 | 6324* 19.97% | 456* 18.07% | 7022* 19.89% | 228 |
| 5 | 5417 | 342 | 4032 | 152 |
| 6 | 3364 | 249 | 3397 | 136 |
| 7 | 2004 | 144 | 2425 | 141 |
| 8 | 802 | 41 | 1695 | 122 |
| 9 | 349 | 29 | 1119 | 107 |
| 10 | 182 | 9 | 762 | 91 |
| Latin 3-grams | nal 3027 | alm 57 | | 3276 127 |
| Latin 4-grams | fiyal 1479 | allah 27 | | 635 51 |

will map the given text to the sequence of the numbers computed from the text D .

The sequence S_l of letters frequencies:

- $D = \langle s_1 s_2 \dots s_{|D|} \rangle$; $s_i \in \Gamma$; - the text as a finite sequence of symbols; $|D|$ – the number of letters in the text D ;
- $\#o_D(s_i)$ – the number of occurencies s_i in the text D ;
- The sequence S_l is prepared by the formula:

$$f_l(s_i) = \frac{\#o_D(s_i)}{|D|}; 1 \leq i \leq |D|.$$

The sequence S_{ln} of $n - gram_l$ frequencies:

- The substring $s_{i+1} s_{i+2} \dots s_{i+n}$, $i = 1, \dots, |D| - n$, $n = 1, 2, \dots$ is called $n - gram_l$ of letters; n - the length of the $n - gram_l$;

- $\#o_{Dn}(l^g)$ – the number of occurrences of n – gram $_l$ $^n g$ in the text D ;
- The sequence S_{ln} contains the frequencies of l^g in the text D

$$f_{ln}(n^g) = \frac{\#o_{Dn}(n^g)}{|D| - n + 1}.$$

3.2 Sequencies of words and n – gram $_w$ s of words

The sequence S_w of words frequencies:

- $D = \langle d_1, \dots, d_N \rangle$; $d_i \in V$ - the text as a finite sequence of words; N - the number of words in the text D ;
- $\#o_w(d_i)$ – the number of occurencies d_i in the text D ;
- The sequence S_w is prepared using the following formula:

$$f_w(d_i) = \frac{\#o_w(d_i)}{N}; 1 \leq i \leq N.$$

The sequence S_{wn} of n – gram $_w$ frequencies:

- The subsequence $d_{i+1}d_{i+2} \dots d_{i+n}$, $i = 1, \dots, N - n$, $n = 1, 2, \dots$ is called n – gram $_w$ of words;
- $\#o_{wn}(n^g)$ – the number of occurrences of n – gram $_w$ $^n g$ in the text D ;
- The sequence S_{wn} is prepared by formula:

$$f_{wn}(w^g) = \frac{\#o_{wn}(w^g)}{N - n + 1}.$$

4 System for anomaly detections

The system is based on an unsupervised learning, the structure is in Fig. 1.

4.1 Desription of the system structure

In the first layer, it has $SOM_x, x \in \{\text{letters, ln-grams, words, wn-grams}\}$ neural networks they are trained to different sequences built according to the text D . The training of each SOM_x is done on sequencies S_l, S_{ln}, S_w, S_{wn} .

The window W is moved through the sequence $S_x, x \in \{l, ln, w, wn\}$ from the beginning to end. $|W|$ is the length of window, m is the length of a moving, both are selective parameters and they are the same in all trained sequencies. We will use $m \in \{|W|/2, |W|\}$, in this case there exist some context between neighbour windows. The dimension of input for each SOM_x is the length of the window W . The number S of input

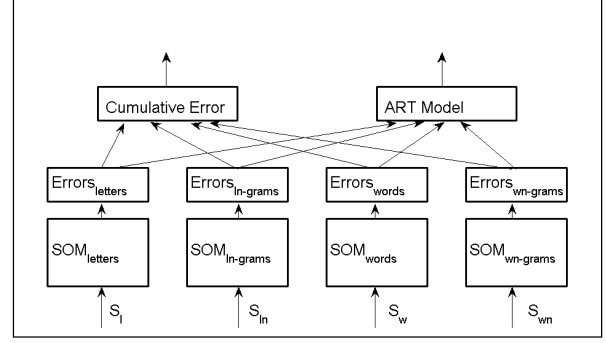


Fig. 1: System for anomalies detections.

vectors is given by $|W|, m$ and by the length of the text $|D|$, $R = \lceil |D|/m \rceil$. If it is necessary the last vector is filled by 0s. The number of iterations It is the next parameter for the training.

4.2 Description of the system computation

After the SOM_x was trained it is possible to evaluate how good was the training prepared by an evaluation of errors for all input vectors (all windows in the sequence). We will use a quantization error Er_x defined by (3) as a measure of proximity input vector \mathbf{x}^+ to the learned winner vector \mathbf{w}_{i^*} of i^* -th neuron (winner for input vector \mathbf{x}^+) in the SOM_x .

$$Er_x(\mathbf{x}^+, \mathbf{w}_{i^*}) = \|\mathbf{x}^+ - \mathbf{w}_{i^*}\|, \quad (3)$$

Using formula (3) it is possible to compute the vectors of quantization errors

$$\{Er_x(\mathbf{x}^+(t), \mathbf{w}_{i^*}(t))\}_{t=1}^R, \quad (4)$$

where R is the number of training vectors, t is the order of the member in input sequence. For the anomaly detections we will use thresholds developed by Barreto a Aguayo (2009). Let α be a significance level ($\alpha = 0.01$ or $\alpha = 0.05$). We suppose the percentage of normal values of the quantization error will be $100 * (1 - \alpha)$. Let N_α be the real number such that a percentage $100 * (1 - \alpha)$ of the error values is less than or equal to N_α . Then

- Lower limit: $\lambda^- = N_{1-\alpha/2}$
- Upper limit: $\lambda^+ = N_{\alpha/2}$

The important interval is $\langle \lambda^-, \lambda^+ \rangle$, the values out of it could be detected as anomalies.

The quantization vectors Er_x are computed in the panels $Errors_x, x \in \{\text{letters, ln-grams, words, wn-grams}\}$ and they are used in two the following evaluations:

- Cumulative Error

$$CEr = \alpha_1 * Er_l + \alpha_2 * Er_{ln} + \alpha_3 * Er_w + \alpha_4 * Er_{wn}, \quad (5)$$

where $\alpha_i, i = 1, 2, 3, 4, \sum_{i=1}^4 \alpha_i = 1$ are parameters for a contribution of Er_i to the cumulative error. The values of the parameters α_i should be chosen after the analysis of all errors.

• ART Evaluation

The layer of SOM networks evaluates the text from more points of view (in the presented model from 4 points of view). It means, after training of all SOM networks we can start to join these results. The used text is divided into the same windows and given to input. The results of winner's neurons can be used to the following analysis. We used ART1 model of neural network in the analysis. ART1 model constructs clusters of text windows. The clusters with more windows explain that text parts are similar (we do not follow copies). In the text parts were used similar words (letters, n -grams of letters and words).

5 Evaluation

We will illustrate the results of our model for two different text in both languages, two of analysed texts (one from each language) are combined from two different texts, and a position of a combination is known too.

5.1 Cumulative Error

We can follow the cumulative errors of two texts combined from two texts written by different authors. In the Figures 2 and 3 a blue vertical lines show the points of the combination.



Fig. 2: Anomalies detections in four sequences for English text E777

In each panel, we can follow a different structure of the errors, mainly in the windows closed to the combination point.



Fig. 3: Anomalies detection in four sequences for Arabic text A14

In the panels of the cumulative error in Figure 2 and 3, there are shown some anomalies over the cyan line. It means the text needs more analysis.

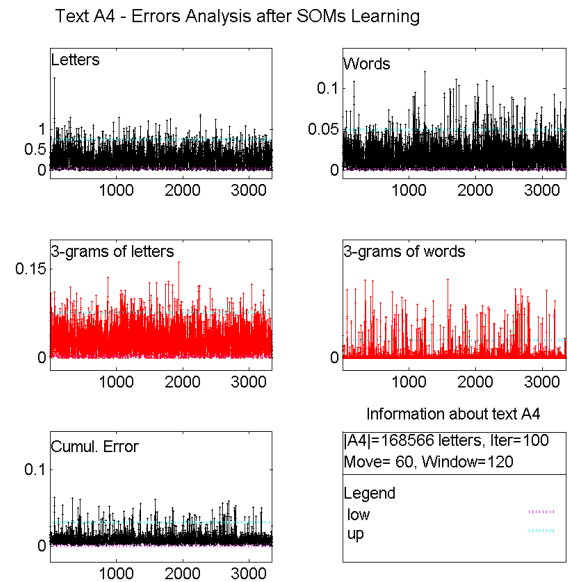


Fig. 4: Anomalies detection in four sequences for Arabic text A4

The long Arabic text written by one author is analysed in Figure 4. The shape of cumulative errors is very balanced. But in the word and 3-grams of words we can

see big errors in some parts of the text. The text needs some next analysis.

5.2 ART Evaluation

The text E777 was followed using 205 windows, numbered according to moving throu the text from left to right. The clustering of windows is the Table 2 and its visualisation is in Fig. 5.

Tab. 2: The clustering of the text E777.

| The number of cluster P1:P2 | The numbers of followed windows |
|-------------------------------|--|
| 0, [3:9] | [0, 85, 90, 118, 123, 139, 141, 169, 175, 190, 196, 199] |
| 1, [2:4] | [1, 87, 115, 143, 167, 177] |
| 3, [3:6] | [3, 8, 29, 61, 110, 114, 150, 178, 181,197] |
| 4, [2:2] | [4, 11, 113, 120] |
| 5, [2:1] | [6, 81, 155] |
| 6, [3:2] | [7, 14, 32, 160, 206] |
| 7, [6:0] | [9, 18, 46, 83, 92, 94] |
| 8, [4:1] | [10, 23, 72, 91, 122] |
| 12, [2:1] | [16, 41, 103] |
| 14, [2:2] | [19, 22, 136, 172] |
| 18, [5:6] | [25, 49, 67, 75, 78, 100, 102, 128, 163, 183, 193] |
| 20, [4:0] | [27, 54, 66, 68] |
| 22, [3:0] | [30, 44, 80] |
| 23[2:2] | [31, 40, 146, 200] |
| 28, [3:14] | [39, 56, 79, 99, 104, 107, 111, 112, 124, 148, 162, 171, 184, 185, 192, 194, 205] |
| 35,[4:0] | [53, 59, 73, 93] |
| 38, [2:2] | [65, 88, 159, 164] |
| 44, [1:2] | [82, 116, 188] |
| 45,[1:2] | [84, 109, 170] |
| 58[0:3] | [132, 157, 168] |
| Clusters with one window, 27 | 15, 33, 34, 37, 40, 48, 49, 50, 53, 54, 56, 59, 60, 61, 62, 64, 65, 66, 67, 71, 72, 73, 74, 75, 76, 77, 79 |
| Clusters with two windows, 33 | 2, 9, 10, 11, 13, 16, 17, 19, 21, 24, 25, 26, 30, 31, 32, 36, 39, 41, 42, 43, 46, 47, 51, 27, 29, 52, 55, 57, 63, 68, 69, 70, 78 |

The analysis of clusters from the point of positions their windows (the first or the second part of the text) show us that there exist clusters with all windows from one part only. These clusters characterise the one part only. If in the text there exist many of them it should be some argument to suppose that text has two parts with different properties. The situation is shown in the Fig 5. For example, the windows in the cluster #28 belong to the second part mainly. In the analysis, there exist more clusters with this property, it means the analysed

text should be composed from two parts (really it is).

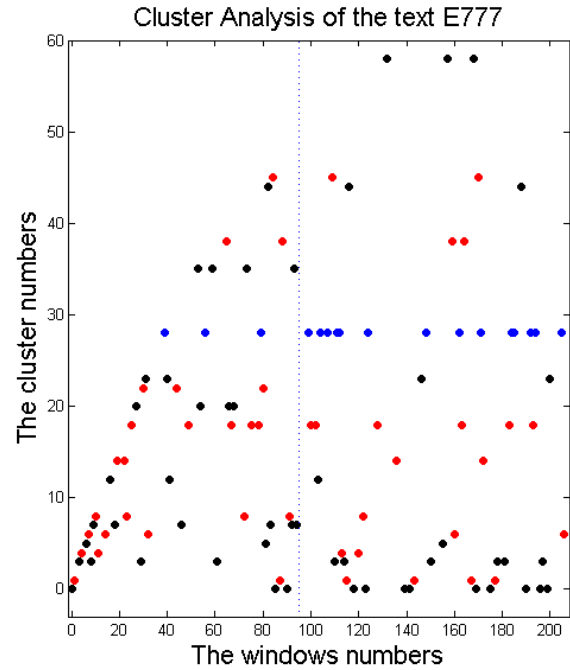


Fig. 5: Visualisation of clusters in the text E777. The biggest cluster is 28 (blue).

The text A14 was followed using 97 windows, numbered according to moving throu the text from left to right. Some clusters belongs to one part only. It mean, the text is probably composed. The clustering of windows is the Table 3, a visualisation of clusters is in Fig. 6.

Tab. 3: The clustering of the text A14.

| The number of cluster P1:P2 | The numbers of followed windows |
|------------------------------|---|
| 0, [11:16] | [0, 10, 15, 19, 20, 34, 36, 38, 50, 54, 59, 63, 67, 77, 82, 83, 84, 85, 86, 87, 89, 90, 91, 93, 94, 95, 97] |
| 1, [7:4] | [1, 11, 21, 23, 37, 39, 47, 61, 62, 88, 96] |
| 10, [6:3] | [12, 13, 14, 16, 22, 60, 64, 80, 81] |
| 11, [8:2] | [17, 41, 42, 46, 52, 55, 56, 57, 71, 72]] |
| 12, [4:0] | [18, 43, 44, 45] |
| 21,[7:1] | [35, 40, 48, 49, 51, 53, 58, 92] |
| 26,[0:3] | [73, 78, 79] |
| Clusters with one window, 18 | 2, 3, 4, 5, 6, 8, 9, 13, 14, 15,17, 18, 20, 22, 23, 24, 25, 28 |
| Clusters with two windows, 4 | 7, 16, 19, 27 |

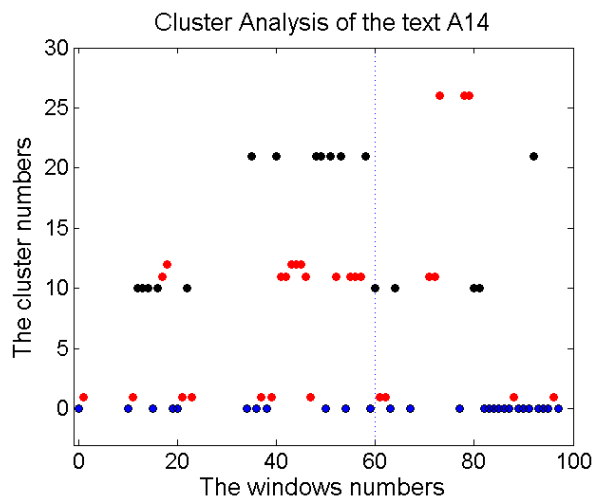


Fig. 6: Visualisation of clusters in the text A14. The biggest cluster is 0 (blue).

6 Conclusion

In the paper we developed the system for anomalies detections in some text. The system was tested on the small number of English and Arabic texts and it is capable to cover anomalies of two texts combinations. But it is necessary to do many statistic tests to its evaluation.

Acknowledgment

The research is supported by the Slovak Scientific Grant Agency VEGA, Grant No. 1/0142/15.

References

- Almarimi, A. and Andrejková, G. (2015a). Discrepancies detection in arabic and english documents. *AC-SIJ Advances in Computer Science: an International Journal*, 4(2322-5157):69–75.
- Almarimi, A. and Andrejková, G. (2015b). Document verification using n-grams and histograms of words. *IEEE 13th International Scientific Conference on Informatics*, (2322-5157):21–26.
- Barreto, G. A. and Aguayo, L. (2009). Time series clustering for anomaly detection: Using competitive neural networks. *Proceedings WSOM 2009, LNCS(5629)*: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*, str. 53–58.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1):1–22.

CorpusArabic (2011). King saud university corpus of classical arabic. <http://ksucorpus.ksu.edu.sa> <http://ksucorpus.ksu.edu.sa>

CorpusEnglish (2011). pan-plagiarism-corpus-2011, <http://www.uniweimar.de/en/media/chairs/webis/corpora/pan-pc-11/>.

Durgin, N. and Zhang, P. (2005). Profile-based adaptive anomaly detection for network security. *SAND2005 (7293)*, str. 1–44.

Hammer, B., Micheli, A., Neubauer, N., Sperduti, A. and Strickert, M. (2005). Self-organizing maps for time series. *WSOM 2005, Paris*, str. 1–8.

Hassan, F. I. H. and Chaurasia, M. A. (2012). N-gram based text author verification. *IACSIT press, Singapore*, 36:67–71.

Jurafsky, D. and Martin, J. H. (2000). *Speech and Language Processing*. Prentice-Hall, 1. vyd.

Kohonen, T. (2007). *Self Organizing Maps*. Prentice-Hall, 2. vyd.

Neme, A., Pulido, J. R., Muñoz, A., Hernández, S. and Dey, T. (2015). Stylistics analysis and authorship attribution algorithms based on self-organizing maps. *Neurocomputing*, str. 147–159.

Sarasamma, S. T. and Zhu, Q. A. (2006). Min-max hyperellipsoidal clustering for anomaly detection in network security. *IEEE Trans. on Systems, Man and Cybernetics - Part B: Cybernetics*, 36(4).

Stamatatos, E. (2006a). Authorship attribution based on feature set subsampling ensembles. *International Journal on Artificial Intelligence Tools*, str. 1–16.

Stamatatos, E. (2006b). Ensemble-based author identification using character n-grams. *In Proceedings of the 3rd International Workshop on Text-based Information Retrieval*, 36:41–46.

Stamatatos, E. (2010). A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, str. 538–556.

zu Eissen, S. M., Stein, B. and Kulig, M. (2006). Plagiarism detection without reference collections. *In: Decker, R., Lenz, H.J. (eds.) GfKI. Studies in Classification, Data Analysis, and Knowledge Organization, Springer Berlin*, str. 359–366.

The temporal profile of contextual plasticity in sound localization

Gabriela Andrejková and Norbert Kopčo

Institute of Computer Science
Pavol Jozef Šafárik University in Košice
Jesenná 5, 04001 Košice
Norbert.Kopco@upjs.sk, Gabriela.Andrejko@upjs.sk

Abstract

This study examines the temporal properties of a new form of auditory spatial plasticity called contextual plasticity. The contextual plasticity is a result of adaptation on the time scale of tens of seconds to minutes, but also shows changes at the scale of seconds. The study analyzes behavioral data from an experimental study in which horizontal-plane sound localization was examined in anechoic and reverberant environments (Kopčo et al., 2005). The results suggest that contextual plasticity has a bi-stable component that is strongly adaptive on the time scale of seconds, while overall the effect is building up over tens of seconds/minutes (Kopčo et al., 2007).

1 Introduction

The problem of sound localization in some context of other sounds is still analyzed from the point of acoustic scene. One of the factors that can influence the listener's ability to localize a target sound presented along with other sounds is perceptual similarity of the stimuli (Best et al., 2008) and temporal distribution of the stimuli. In the (Kopčo et al., 2007), it is described a special type (1-click) of stimuli preceding the target sound on the time scale of seconds to tens of seconds. This spatial auditory plasticity is called a contextual plasticity.

Effects of preceding stimuli on a target localization can be described by various candidate mechanisms operating on time scales from milliseconds to ten of seconds. For example, a precedence effect (Brown et al., 2015), an adaptation of brainstem neural representation due to prolonged exposure (the effects of a stationary adapting noise stimulus on the subsequent auditory localization in the vicinity of the adapting stimulus, 4 min of continuous noise at the start of each block of trials) to the preceding stimuli (Carlile et al., 2001). Best et al., (2008) showed that sequential grouping cues (i.e. effects of preceding distractor types) can have influence on whether binaural interference occurs.

In the current study, click stimuli were used as targets. On some trials they were preceded either by a grouping (1-click) or a streaming (8-click) distractor, coming from the frontal and lateral positions relative to the subject. The contextual plasticity was evaluated as a perceptual bias, computed as the difference between the average target-alone responses in frontal- vs. lateral-distractor runs. The observed bias was up to 10°.

Here, temporal analysis was done by dividing each experimental run with approximate duration of 5 minutes into four equal-duration subruns. When 1-click distractor was used in the contextual trials (a trial when an identical sound target is presented from a random location following after an identical distractor coming from a fixed, a priori known location), the contextual bias built up within first three subruns, reaching magnitude of approx. 7° in classroom, and 10° in anechoic room. When context consisted of both 1-click-distractor and 8-click distractor trials, the contextual bias did not depend on type of the room and its buildup continued up to the last, 4th subrun, where it reached 12°-13°. Separating this bias according to whether it was preceded by 1-click-distractor contextual trial or 8-click-distractor contextual trial showed that the bias built up differently for the two conditions. Specifically, in the 2nd subrun, when target-alone trial target is preceded by 8-click-distractor trial, the bias is approx. 5 degrees larger than when preceded by 1-click-distractor trial.

In this paper, we will first review the results from a previous study (Kopčo et al., 2007). These will be referred to as Exp. 1. The results will be compared to new results from the experiment referred to as Exp 2. The effect of the preceding distractor to biases in Exp 2 was analyzed in (Kopčo et al., 2016).

2 Methods

2.1 Subjects

Seven subjects participated in a classroom and four of them participated in an anechoic room. All subjects had

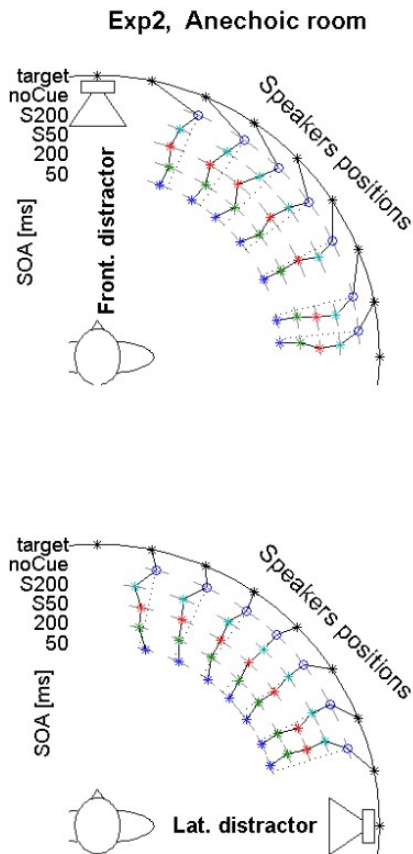


Fig. 2. Mean localization responses in the classroom and in the anechoic room separated according to the frontal and the lateral distractor. Each panel shows the across-subject mean and standard error in perceived target lateral angle as a function of actual target lateral angle for different ISI.

For each subject, 32 responses were collected in total for each combination of (4 ISI + no distractor), two distractors and seven target lateral angles.

The first analysis is shown in the Fig 2. It shows the raw data, plotting the across-subject mean and standard error in the perceived target location as a function of the actual source lateral angle for frontal and lateral distractor in the classroom and in the anechoic room.

The black asterisks are drawn in the actual angles of the target loudspeakers. The results are drawn in the parts of circles according to the legend (no Cue, S200 - ISI 200 ms, 8-click distractor, S50 - ISI 50 ms, 8-click distractor, 200 - ISI 200 ms, 1-click distractor, 50 - ISI 50 ms, 1-click distractor).

The analysis of the biases shows that for the data of the frontal distractor, localization responses are biased towards the side. For the data of lateral distractor, the most of localization responses are biased towards the midline. The detailed analysis of distractor effects can be found in (Kopco et al., 2016).

4 Temporal Profile of Plasticity

Kopčo et al. (2007) presented an analysis of the temporal profile of the contextual effect of 1-click distractor. The results are presented using the across-subject mean in the contextual difference as a function of subruns. The experimental runs were subdivided into 4 subruns for this analysis. The main results are (Fig. 3A):

Results in the anechoic room: The contextual bias of no distractor trials built up over the four subruns, growing from roughly 6 to 9°.

Results in the classroom: The contextual effect increased with subruns, it grew from 4 to 8° across the four subruns for no distractor trials.

The new experiment follows more complex auditory scene - trials using two types of distractor, 1-click and 8-click.

4.1 1-click versus (1-click + 8-click)

Fig. 3, panel A contains temporal profile of the contextual bias for the experiment Exp 1 described in (Kopčo et al., 2007) and of the new experiment Exp 2.

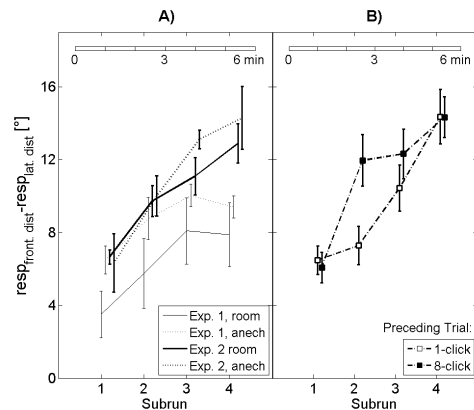


Fig. 3. The contextual bias - across lateral angle average difference in responses as a function of subruns within experimental runs (no distractor trials). Panel A) contains the results of Exp1 and Exp2. Panel B) shows contextual bias for 1-click and 8-click distractor trials in Exp2.

When 1-click distractor was used in the contextual trials in experiment Exp1 (thin lines in the panel A, the contextual bias increases within first three subruns and it was reached 7° approximately in the classroom and 10° in the anechoic room.

In the experiment Exp 2 the context was mixed using 1-click and 8-click distractor trials. The contextual bias here (thick lines in the panel A) was increasing within four subruns and it does not depend on the type of the

room. The maximum magnitude was approximately 12-13°. It is possible to suppose that the magnitude should increase in prolonged subruns. The results of the analysis of variance (ANOVA) are in the Tab. 1.

| Factor | df | F Significant |
|------------------|--------|---------------|
| Experiment (Exp) | 1, 3 | 15.36 ** |
| Room (R) | 1, 3 | |
| Subrun (S) | 3, 9 | 32.16 *** |
| Target (T) | 6, 18 | |
| Ex x Room | 1, 3 | |
| Ex x Target | 6, 18 | |
| Ex x Subrun | 3, 9 | |
| Room x Subrun | 3, 9 | |
| Room x Target | 6, 18 | |
| Target x Subrun | 18, 54 | |
| Exp x R x T | 6, 18 | |
| Exp x R x S | 3, 9 | 5.03 ** |
| Exp x T x S | 18, 54 | |
| R x T x S | 18, 54 | |
| Exp x R x T x S | 18, 54 | |

Significance levels are as follows: ** $p < 0.001$ and *** $p < 0.005$.

Tab. 1: Results of four-way repeated measures ANOVA performed on the response bias data in both experiments and both rooms for targets and subruns. F values and significance values are listed only for effects and interactions with significance level $p < 0.05$.

Separating this bias according to a conditions in a preceding trial, especially 1-click or 8-click distractor trial, Fig. 3, panel B, showed that the bias increases in a different way for them. It is possible to follow it from the 2nd subrun, when the no distractor target trail is preceded by 8-click distractor trial has the bias approximately 5° larger than preceded 1-click distractor trials. The results of two-way repeated measures ANOVA are in the Tab. 2.

| Factor | df | F Significant |
|----------------------|-------|---------------|
| Distractor type (DT) | 1, 6 | 7.23 ** |
| Subrun | 3, 18 | 22.98 *** |
| DT x Subrun | 3, 18 | 4.67 ** |

Significance levels are as follows: ** $p < 0.001$ and *** $p < 0.005$.

Tab. 2: Results 2-way ANOVA applied to subruns and the type of distractors (1-click and 8-click).

4.2 Context biases in rooms

In Fig. 4, it is shown a comparison of the contextual bias (analyzed as a difference of the frontal and lateral distractor responses in trials) in the anechoic room and in the classroom separately.

Results in the anechoic room: The contextual biases of all three trial types are increasing in the second subrun, approximately in 4° for no distractor trials and 2° for 1-click and 8-click distractor trials. In the third subrun the increasing continues for no and 1-click distractor trials, it is larger for no distractor trials, the contextual bias of 8-click distractor trials show a stable value. The differences among contextual biases in the fourth subrun are very close to the differences in the first subrun, but a different trend is followed in the second and the third subrun. The results of two-way repeated measures ANOVA are in the Tab. 3.

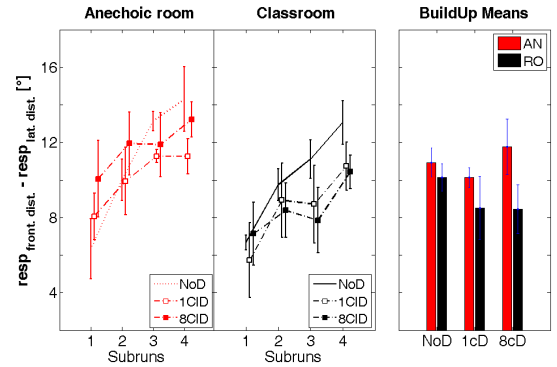


Fig. 4. The contextual bias analyzed in the rooms. In both rooms, it is shown contextual bias for no distractor, 1-click, and 8-click distractor trails. The last panel shows mean values and standard errors for all responses in the anechoic room (the red columns) and the classroom (the black columns).

| Anechoic room | | |
|---|-------|---------------|
| Factor | df | F Significant |
| Distractor type (DT) - no distractor, 1-click, 8-click distractor | 2, 6 | |
| Subrun | 3, 9 | 10.34 *** |
| DT x Subrun | 6, 18 | |

Significance levels are as follows: *** $p < 0.005$

Tab. 3: Results 2-way ANOVA applied to subruns and the type of distractors (no distractor, 1-click and 8-click) in the anechoic room.

Results in the classroom: The contextual effect of no distractor trials has an increasing trend, approximately 4.5°, but it is not as large as in the anechoic room. The contextual effect trends of 1-click and 8-click distractor trials are very similar. The values in the fourth subrun are very similar. The results of two-way repeated measures ANOVA are in the Tab. 4.

The comparison of results in both rooms for across-subrun-average values of all three trial types is in Fig. 4, the third panel. The 8-click distractor trials show the largest difference of the contextual bias means in

rooms. The result of the contextual bias trends together in both rooms is shown in Fig. 3, panel B. The influence of 8-click distractor trials in the contextual bias is larger than the influence of 1-click distractor trials in the second and in the third subrun. The fourth subrun presents that using different build up the contextual effect finish with very similar values.

| Classroom | | |
|---|-----------|----------------------|
| Factor | df | F Significant |
| Distractor type (DT) - no distractor, 1-click, 8-click distractor | 2, 12 | |
| Subrun | 3, 18 | 11.37 *** |
| DT x Subrun | 6, 36 | |

Significance levels are as follows: *** $p < 0.005$

Tab. 4: Results 2-way ANOVA applied to subruns and the type of distractors (no distractor, 1-click and 8-click) in the classroom.

5 Discussion

Kopčo et al. (2007) showed that the contextual bias occurs in trials with and without distractor. It causes a shift in the perceived locations of the targets for both distractors. This bias is systematically away from the distractor position in the run and builds up over the course of minutes. In the anechoic room, it is possible to observe that the lateral distractor suppresses the spatial percept of lateral target.

Experiment Exp 2 showed that the contextual effect is stronger for the trials consisted from 1-click and 8-click distractor in comparison to Exp 1 (Kopčo et al., 2007) with 1-click distractor trials only (see Fig. 3, panel A). It suggests that the contextual effect of the distractor - target trials on no distractor trials responses is more complex than a simple adaptation in the spatial map.

6 Conclusion

The contextual bias induced by 8-click distractor is stronger than the contextual bias induced by 1-click distractor and it is larger in the classroom than in the anechoic room. This observation is consistent with a bottom up adaptation mechanism sensitive to the distribution stimuli (Kopčo et al., 2015).

Acknowledgement

This contribution was supported by Slovak Research and Development Agency under grant project APVV-0452-12 Spatial Attention and Listening in Complex Environments.

References

- Best, V., Gallun, F. J., Carlile, S., Shinn-Cunningham, B. G. "Binaural interference and auditory grouping." *J. Acoust. Soc. Am.* 121(2) (2007): 1070-1076.
- Best, V., Ozmeral, E. J., Kopčo, N., Shinn-Cunningham, B. C. "Object continuity enhances selective auditory attention." *Proceedings of the National Academy of Sciences of the USA* (2008): 13174-13178.
- Brown, A. D., Stecker, G. Ch., Tollin, D. J. "The precedence effect in sound localization." *Journal of the Association for Research in Otolaryngology* 16, (2015): 1-28.
- Carlile, S., Hyams, S., and Delaney, S. "Systematic distortions of auditory space perception following prolonged exposure to broadband noise." *J. Acoust. Soc. Am.* 110 (2001): 416-424.
- Kopčo, N., Andrejková, G., Best, V., Shinn-Cunningham, B. G. "Streaming and sound localiyation with a preceding distractor. ." *J. Acoust. Soc. Am.* (2016): submitted.
- Kopčo, N., Best, V., and Shinn-Cunningham, B. G. "Sound localization with a preceding distractor." *J. Acoust. Soc. Am.* 121 (2007): 420-432.
- Kopčo, N., Best, V., Shinn-Cunningham, B. G. "Click versus Click-Click: Influence of a Preceding Stimulus on Sound Localization." *Association for Research in Otolaryngology, Abstract #965* (2005).
- Kopčo, N., Marcinek, L., Tomoriová, B. and Hladek, L. "Contextual plasticity, top-down, and non-auditory factors in sound localisation with a distractor." *J. Acoust. Soc. Am.* 137 (4), EL 281. (2015).

Konceptuální analýza vědomí

Bc. Jiřina Babická

Západočeská univerzita v Plzni
Sedláčkova 268/38
jirinab@students.zcu.cz

Abstrakt

V odborných textech často dochází k významovým posunům při užívání pojmu vědomí. Vzniká tak množství podobných konceptů, které se sémanticky překrývají. Analyzovala jsem tedy koncepty vědomí a zařadila do čtyř konceptuálních tříd 0 - III. Do nulté třídy spadají pojmy fenomenálního vědomí, a třídy I - III obsahují pojmy vědomí, které reflektují míru uvědomění. Rozdělení pojmů do čtyř tříd má posloužit jako metodologický nástroj, který by měl usnadnit orientaci v kognitivněvědním terminologickém prostoru.

1 Analýza

Tato analýza spočívá v rozřídění fenomenologických konceptů vědomí na ty, které reflektují stupně uvědomění, a na ty, které postihují fenomenální aspekt života organismu. Koncepty jsou rozděleny do čtyř tříd.

Protože se v mé analýze jedná spíše o vytvoření metodologického nástroje, který má usnadnit orientaci v terminologii, k úrovním uvědomění jsem přiřadila zvláštní nultou třídu pro koncepty, u kterých se nejedná o míru přístupnosti mentálních obsahů v uvědomování. Nultá třída zahrnuje koncepty, u kterých je kladen důraz na komponentu, popisovanou jako kvalitativní, subjektivní prožívání, "jaké to je být" apod. Tato komponenta je životu organismu vlastní, kromě některých kvalitativních a kvantitativních poruch vědomí.¹ U pojmů této třídy se jedná o specifický způsob existence lidí nebo širšího okruhu živočichů². Je možno ji tedy chápat jako vlastnost, která prostupuje vědomím nehledě na stupně uvědomění. Proto tuto třídu označuji číslem nula.

Třídy I - III odrážejí stupně uvědomování, čili míru zpřístupnění mentálního obsahu vědomému prožívání. Tyto tři třídy se do jisté míry shodují se dvěma klasifikacemi (viz níže), ve kterých jsou stejné koncepty nahlíženy jako stupně uvědomění. Dvě

¹ Například kóma či určitý typ epileptického záchvatu.

² Ačkoli existují také směry, které počítají s elementární fenomenalitou i u neživých entit, jako například panpsychismus či panprotopsychemus.

následující klasifikace jsou vytvořeny za jiným účelem rekonstruování struktury vědomí podle fenomenologických pojmů, ale zde poslouží jako ověření toho, že fenomenologické koncepty i přes svou synonymitu a překrývání je možné utřídřit podle širšího významu.

Autorem první klasifikace je Alain Morin (2006), jenž analyzoval koncepty uvědomění a sebeuvědomění. Na základě této analýzy vymezil škálu pěti úrovní uvědomění. Tato škála sahá od nevědomí po uvědomování si vlastního sebeuvědomění.³

Podobně ke konceptuální heterogenitě přistoupili například Berkovich-Ohanová a Glicksohn (2014). Ti představili tzv. model prostoru vědomých stavů (consciousness state space model). Tento prostor je rozdělen na jádrové a rozšířené vědomí⁴, a na tři psychologické dimenze: a čas, uvědomění a emoce. Autoři přisoudili jádrovému a rozšířenému vědomí charakter domén, ve které se prolínají obsahy uvedených třech dimenzí. Tím se jejich klasifikace liší od Morina, který jádrové a rozšířené vědomí chápe též jako úrovně uvědomění. Při klasifikaci tříd I - III se spíše přikláním k modelu Berkovich-Ohanové a Glicksohna a dichotomii jádrového a rozšířeného vědomí pokládám jako jinou doménu prožívání.

2 Konceptuální třídy

Zde je vhodné zmínit, že si nečiním nárok na analyzování veškerých konceptů, které se na poli kognitivní vědy vyskytují. V následujících podkapitolách je věnována pozornost pouze určitému

³ Do svého rozboru konceptů sebeuvědomění Morin zahrnuje i koncepty jáství, které v jednotlivých třídách uvádím také, ačkoli by se toto zařazení mohlo zdát diskutabilním. Například Neisserových pět úrovní "já" (viz konceptuální třídy) Morin nazývá úrovněmi vědomí (2006, s. 362), přestože Neisser (1988, 1997) o nich jako o úrovních vědomí nehovoří. Domnívám se, že typy "já" v Neisserově pojetí jsou aspekty uvědomění, které mohou vykazovat charakteristiky jednotlivých konceptuálních tříd.

⁴ Autoři přejímají rozdělení na jádrové a rozšířené vědomí od Antonia Damasia (1999).

vzorku konceptů, na kterém jsou konceptuální třídy vystavěny.

Zařazení konceptů do tříd není bez problémů, neboť některé pojmy vykazují rysy, společné dvěma třídám nebo jsou zařaditelné na jejich pomezí. Případné problematické zařazení budu komentovat přímo u jednotlivých pojmů.

2.1 Třída 0

V této třídě jsou obsaženy koncepty, u nichž je zdůrazňována fenomenalita jako aspekt vědomí, který je vlastní životu lidských bytostí, případně jiných živočišných druhů. V souvislosti s tímto aspektem vznikly určité pochybnosti o tom, jestli je možné vysvětlit vznik subjektivního prožívání na materiálním substrátu.⁵ V nahlížení fenomenality je možné spatřovat několik rovin: pohled na fenomenalitu jako vlastnost jak smyslové percepcce, tak i kognice, oproti přesvědčení, že fenomenalita je záležitost smyslové percepcce; dále lze spatřit spory o to, zda má fenomenalita jen vědomou nebo i nevědomou podobu.⁶

Koncept, který představuje veškeré prožívání bez rozdělování kognice a percepcce je "subjektivita", postulovaná Thomasem Nagelem (1974) je intrinsickou vlastností živých bytostí, je zcela nepřenositelná, avšak částečně sdělitelná pomocí jazyka a v rámci téhož druhu. Podobným konceptem je "Bytostní vědomí" ve smyslu, v jakém jej užívá Tim Bayne (2007). Bayne chápe bytostní vědomí jako fenomenálního vědomí v holistickém slova smyslu.⁷

Pojem "Kvália" (Lewis, 1929) sice neoznačuje typ vědomí, ale jeho fenomenální aspekt. Ta mohou být definována jako jednoduché smyslové kvality (Damasio, 1999), jež jsou neredukovatelné na menší části, jsou přístupné pouze subjektu. Stejně jako k fenomenálnímu vědomí lze i zde zaujmout různá stanoviska o povaze kválií. Kvália byla původně nahlížena jako pouze percepční, objevují se však autoři, pro které má kognitivní stránka též kvalitativní charakter (viz např. Shields, 2011). Například Bernard Baars (1997) neuznává kvalitativní povahu abstraktního myšlení apod. (s. 93). Na druhou stranu Baars (1997) přiznává, že kvália mohou být i nevědomá.

"Fenomenální vědomí" Neda Blocka (1995) je koncept z toho názorového proudu, který spojuje fenomenalitu pouze se smyslovou percepcí. Fenomenální vědomí je definované jako smyslová zkušenost, emocionální

⁵ Viz tzv. "těžký problém" (Chalmers, 1995) nebo "explanační mezera" (Levine, 1983)

⁶ Podrobněji tento spor popisuje například Marvan & Polák, 2015, kap. 4.2.

⁷ Naopak například pro Petera Carrutherse nebo Davida Rosenthala je bytostní vědomí pojem pro bdělost.

stavy a myšlenky. Na rozdíl od přístupového vědomí je obsahově bohatší, neboť přístupové dokáže postihnout menší část toho, co je prožíváno. Pro Blocka fenomenální vědomí existuje i v nevědomé podobě.⁸ Fenomenální vědomí jako percepční chápou i Tomáš Marvan a Michal Polák (2015), kteří stejně tak obhajují existenci fenomenálních obsahů v nevědomí (tamtéž).

Dalším pojmem, označujícím smyslovou fenomenalitu je "mikrovědomí" (Zeki, 2003). Teorie mikrovědomí říká, že dílčí fenomenální kvality jsou časově nesynchronizované a jsou zpracovávány neuvědomovaně⁹ v různých částech mozku (např. barva, pohyb). Jejich synchronizací potom vzniká makrovědomí, které již podléhá uvědomování. Zekiho mikrovědomí je případ hraniční pozice. Tím, že se jedná o nevědomé zpracování smyslových vstupů, splňuje kritéria pro zařazení mezi pojmy pro nevědomé procesy v třídě I, avšak akcentováním fenomenální stránky tohoto procesu je relevantní pro nultou třídu.

Podobně na pomezí první a druhé třídy je možno spatřovat tzv. "primární vědomí", které použil Johanson a kol. (2003) ve stejném smyslu jako Edelman (2004). Johanson tvrdí, že primární vědomí je jedinou úrovní vědomí, ve které je obsažena subjektivní zkušenost (2003, s. 280). Primární vědomí tito autoři dělí na centrální a periferní. Centrální zaujímá střed pozornosti a periferní se miní částečně vnímání toho, co se děje mimo centrum pozornosti. Přes důraz na fenomenální povahu tohoto pojmu, by bylo možno jej kvůli jeho obsahu zařadit i do třídy I (u periferního vědomí) a II (u vědomí centrálního).

2.2 Třída I

Zde jsou zahrnuty pojmy, které se pohybují na hranici uvědomování, a je možno je odlišit podle více kritérií. Za prvé je lze dělit podle toho, jestli se jedná o vědomí v rámci vývojového stádia (z hlediska ontogeneze i fylogeneze). Takové vědomí je potom jakýmsi výchozím stavem, ve kterém se vyvíjejí vyšší kognitivní funkce a mentální činnosti, zejména v souvislosti s učením jazyka. Za druhé je možno vymezit skupinu pojmů, které se týkají buď pouze percepcce, nebo popisují zpracování nepercepčních informací na nevědomé úrovni.¹⁰

Z typů, odrážejících některé počáteční vývojové stádium je to například "anoetické vědomí". Endel

⁸ Pro tuto tezi argumentuje výsledky experimentu, prováděného Georgem Sperlingem v 60. letech (Block, 2011)

⁹ Michael Cohen a Daniel Dennett (2011) se staví kriticky vůči oddělování fenomenálního prožitku a přístupového vědomí, protože se tak uzavírá možnost prožitků podrobovat zkoumání

¹⁰ V případě pojmu "minimální vědomí" je nutné jej nespojovat s tzv. stavem minimálního vědomí, který je označením kvantitativní poruchy vědomí.

Tulving definoval tři typy vědomí v návaznosti na typ paměti. Prvním z nich, je vědomí anoetické (1985), neboli nepoznávající¹¹ vědomí. Tento typ je spojen s procedurální pamětí. Nepoznávajícím je v tom smyslu, že organismus dokáže zaznamenávat stimuly pouze pomocí smyslové percepce, které se navíc odehrávají v přítomném čase. Dále by bylo vývojových pojmů možno počítat "minimální vědomí" (Zelazo, 1996), jenž je typické pro novorozenecké období, tudíž je nekonceptuální a nedokáže se orientovat na budoucnost a minulost, vztahuje se pouze k přítomnosti. "Protovědomí"¹² (Hobson, 2009) by bylo možno definovat jako prvotní stav vědomí, který vzniká ve třetím trimestru prenatálního období, a *de facto* je snovou fází spánku. Během života se toto vědomí objevuje už stále jen jako fáze REM. Hobson (2014, 2009) užívá "primární vědomí", jež je typické pro savce, avšak člověk se z tohoto vědomí dokáže posunout do vědomí sekundárního vlivem jazyka (třída II).

Další skupinou jsou pojmy pro neuvědomované stavy. "Minimální vědomí" podle Armstronga (1981) jsou utlumené nebo inaktivované mentální stavy, představy, přesvědčení nebo touhy. Jsou tzv. kauzálně pasivní, a tudíž nejsou schopny zasahovat do jiných mentálních dějů. Minimální vědomí probíhá například během spánku, nebo jiných činností, kdy jsou uvedené mentální stavy a obsahy pozastavené. Pojem "nevědomí" (Schooler, 2002) označuje neuvědomované monitorování vědomí¹³, ovlivňování výběru a dosahování cíle, a utváření obsahu myšlení. "Nevědomá mysl" podle Farthinga (1992) sestává z nepovšimnutých percepčních vjemů, automatických kognitivních, sensorických i motorických procesů. Farthing vymezuje uvnitř nevědomé mysli ještě škálu dalších pochodů: smyslové vstupy, které jsou registrované avšak ne povšimnuté, dále deklarativní dlouhodobou paměť, automatické kognitivní a senzomotorické programy, nevědomé motivy, a nakonec podvědomé znalosti (1992, s. 21). Kategorie nevědomých stavů zahrnuje také tzv. "marginální vědomí" (Gurwitsch, 2010). Tento typ je souhrnem neuvědomovaných stavů těla, tedy percepční a propriocepční procesy, různé potřeby a myšlenky. Nicméně tyto neuvědomované procesy mohou za určitých podmínek vytanout a objevit se v "tématu", tj. v centru pozornosti, nebo v jeho okolí, "tématickém poli". Touto dispozicí vytanout v uvědomění se marginální vědomí blíží konceptu předvědomí. "Předvědomím" chápe Urbano (2014, 2009) pomyslný prostor, v němž se zpracovávají mentální obsahy, které se ještě nedostaly do centra pozornosti.

¹¹ Tulving (1985, s. 3) zde užívá "not-knowing", avšak vzhledem k etymologii slova "anoetický" překládám jako "nepoznávající".

¹² Nejedná se o totéž protovědomí, o kterém se hovoří v protopanpsychismu.

¹³ Jedná se o prostřední z triády vědomí Jonathana Schoolera.

Z hlediska částečného uvědomování existuje pojem "okrajové vědomí" (James, 1950). Okrajové vědomí je nepercepční pocit¹⁴, který doprovází kognitivní operace. Jedná se o intuici, či pocit správnosti. K tomuto konceptu se v současnosti někteří autoři vracejí (např. Vandekerckhove & Panksepp, 2009; Mangan, 2001). Podobnou charakteristiku nese Farthingovo "periferní vědomí", které se nachází mezi jeho "nevědomím" a "vědomím". To označuje mentální obsahy, vyskytující se na okraji tzv. fokální pozornosti (Farthing, 1992). Těmto zpola uvědomovaným typům odpovídá i tzv. "tématické pole" (Gurwitsch, 2010, s. 310), které ulpívá kolem "tématu" (viz třída II.) a utváří jeho kontext. V tématickém poli dochází k jakémusi částečnému uvědomování, simultánně se soustředováním pozornosti k obsahu v tématickém centru. Tématické pole tvoří s marginálním vědomím tzv. vědomí periferní.

"Ekologickým já" Neisser (1988) pojmenovává uvědomování si vlastní tělesné kondice a okolního prostředí, které na tělo působí. Toto vědomí je spíše percepční, nekonceptuální, a je možné ho chápat na pomezí třídy I a II.

2.3 Třída II

Tato třída obsahuje pojmy pro uvědomění v takovém rozsahu, kdy je osoba schopná plně reflektovat v mysli dění v okolním prostředí. Zde je možné rozlišit koncepty pro takovou úroveň uvědomění, u něž není kladen důraz na jazyk s pamětí, oproti tomu je podstatná schopnost plně si uvědomovat smyslové vjemy. Druhým typem je úroveň, na které je podstatným rysem zvnitřnění jazyka, které již umožňuje manipulaci s pamětí. První případ těch spíše percepčních konceptů se blíží charakteristice třídy I. Ten druhý je naopak blíže třídě III, protože znalost řeči implikuje i znalost slov, díky kterým subjekt může odlišit vlastní osobu od ostatních apod.

"Percepční vědomí" (Armstrong, 1981, s. 723) znamená stav, kdy subjekt plně vnímá své okolí pomocí smyslů (včetně iluzí), a to během doby, kdy je subjekt bdělý. Toto vědomí je orientováno pouze na přítomnost, vzhledem k tomu, že se zakládá na smyslovém prožívání momentální situace. Pojem "Primární vědomí" (Farthing, 1992) zahrnuje také přímou uvědomovanou percepci, myšlenky, pocity a vzpomínky, nebo vnitřní řeč. Jeho vznik začíná již v rané fázi dětství a jeho další aspekty se vyvíjí v později. U pojmů v tomto odstavci si lze všimnout podobnosti s Blockovým fenomenálním vědomím. Rozdíl je pouze ve způsobu pojmenování.

¹⁴ Tento pocit je Jamesem popisován jako psychická příchut' nebo nádech.

Na předchozí pojmovou skupinu navazuje "noetické vědomí" (Tulving, 1985), které se soustředí k přítomnosti, je percepční, a podmínkou pro jeho vznik je bdělost a znalost jazyka, protože se pojí se sémantickou pamětí. Čistě jazykový typ uvědomění je "rekurzivní vědomí" (Zelazo, 2004), jenž je téměř závislé pouze na jazyku, takže v ontogenezi nastupuje později v průběhu učení, a to po prvním roce života. V tomto vědomí je jedinec schopen přiřazovat sémantické kategorie k percepční zkušenosti.

Do třídy II jsem zařadila následující koncepty, které svým rozsahem hraničí s třídou III, protože připouštějí i možnost vzniku sebereflektivní mentální činnosti. Jedním takovým je tzv. "téma" (Gurwitsch, 2010), které lze popsat centrum pozornosti, ať již zaměřené na okolí nebo vlastní psychické nitro. V této podskupině spadá i známé "přístupové vědomí" (Block, 1995). To má zcela informační charakter, je tvořeno reprezentacemi fenomenálního obsahu, které hrají roli v procesech jako odůvodňování a kontrola jednání a produkce řeči. Tyto procesy jsou taktéž součástí přístupového vědomí. Za hraniční typ je možné považovat sjednocené vědomí (Zeki, 2003), jež vzniká dostatečným integrováním makrovědomí. Ve sjednoceném vědomí se pravděpodobně odehrávají vyšší kognitivní procesy včetně reflektivních a sebereflektivních stavů.¹⁵ Patří tedy do třídy II a III.

Mezní kategorii v případě jáství lze považovat "interpersonální já", které znamená, že si subjekt plně uvědomuje dění v okolním prostředí. Zahrnuje vnímání interakcí mezi subjektem a dalšími lidmi (Neisser, 1988). Morin tento typ zařazuje již do sebeuvědomění, avšak vzhledem k tomu, že subjekt nevnímá pouze sebe, ale i své okolí a další osoby v něm, ponechávám interpersonální já zde.

2.4 Třída III

V této třídě jsou zahrnuty koncepty schopnosti zaměřit se na své vlastní mentální stavy a na sebe sama. Všechny tyto pojmy vyžadují osvojenou řeč, což jde ruku v ruce s ontogeneticky pozdním nástupem sebereflektivních stavů.¹⁶ Lze zde rozlišit rozdílné stupně uvědomění: uvědomování nebo všímání si sebe sama v kontextu okolí, uvědomování si obsahu vlastních myšlenek, a jako nejvyšší úroveň je definována schopnost uvědomit si vlastní sebereflexi, někdy též metauvědomování¹⁷. Někteří autoři jako například Endel Tulving, David Armstrong nebo Philip D. Zelazo spojují sebereflektivní úroveň uvědomění s

paměťovými mechanismy a s chápáním koncepce minulosti a budoucnosti.

Mezi koncepty uvědomění, které vzniká v souvislosti s učením se jazyku, patří "autonoetické vědomí" (Tulving, 1985). To tedy také vzniká v pozdějším věku. Umožňuje vnímat osobní čas, a je nezbytné pro pamatování si událostí. Je spjata s epizodickou pamětí, díky čemuž je osoba schopná vyvolat vzpomínky, které mají podle Tulvinga zvláštní charakter, a totiž že vzpomínky přinášejí pocit existence sebe sama v minulosti. Tento popis nápadně připomíná následující Armstrongovo vědomí introspektivní. "Introspektivní vědomí" (Armstrong, 1981), umožňuje reflexi vlastních mentálních stavů, Armstrong ho považuje za vnitřní percepci. Introspektivní vědomí tím, že zkušenost vztahuje k vlastní osobě, umožňuje vstup informací do paměti. Podobné charakteristiky má i "reflektivní vědomí" (Farthing, 1992). To je definováno jako dovednost zaměřit se na vlastní mentální stavy, jejichž obsahem jsou jevy, události a věci z okolního prostředí. Součástí reflektivního vědomí je introspekce, jež má dosah do primárního vědomí. Farthing považuje toto vědomí jako nezbytný základ pro "sebeuvědomění", které v jeho teorii znamená chápání sebe sama jako jedinečné osoby, která se odlišuje od druhých lidí, má své vlastní historii a budoucnost. Zelazo (2004) vymezil "vědomí sebe sama"¹⁸ a "reflektivní vědomí", jež jsou také vývojovými stupni, které souvisí s osvojováním pojmů, rozpoznáváním své vlastní osoby, a vytvářením si teorie mysli. Závislejší na jazyce je též "sekundární vědomí" (Hobson, 2009). Jedná se o uvědomění, které zahrnuje percepci a emoce, obohacené o abstraktní analýzu (myšlení) a metakognitivní komponenty vědomí (uvědomování si uvědomění).

"Metavědomí" (Schooler, 2002) je označení pro explicitní nahlížení vlastních mentálních stavů, slovy Schoolera, dochází k re-reprezentování¹⁹ vlastních stavů vědomí. Schooler zdůrazňuje, že metavědomí není nepetržitě, ale vyvstává na střední úrovni vědomí (tř. II.) a zároveň může být přerušováno některými jinými mentálními činnostmi. U Hobsona (2009) je metavědomí schopností vnímat sebe sama jako původce jednání. Toto metavědomí se objevuje i v jevu, zvaném lucidní snění. Hobson v případě metavědomí a sebereflektivního uvědomění též hovoří o významném vlivu evoluce řeči. Například tvrdí, že lidé mají na rozdíl od ostatních savců rozvinuté zakoušení bdělosti díky sebezpozorování spolu se schopností se verbálně vyjádřit o svých stavech (2014, s. 77).

¹⁵ Zeki se však vznikem ani obsahem tohoto vědomí podrobněji nezabývá.

¹⁶ Zde je nutné odlišit koncepty III. třídy od Damasiova prelingvistického jádrového jáství, které by se blížilo spíše třídě I zejména pro svůj prelingvistický charakter.

¹⁷ Podrobněji například v: Morin, 2006.

¹⁸ Originální pojem je "self-consciousness". Záměrně zde nepřekládám jako sebeuvědomění, abych

¹⁹ Tento neobvyklý slovní tvar ponechávám podle "re-representation", užitého autorem (Schooler, 2002).

Pokud se týká konceptualizace jáství, se třemi typy seberefektivní úrovně pracuje Neisser (1988). "Rozšířené já" odkazuje na rozšířené pojmání času - osoba si dokáže promítnout sama sebe do budoucnosti a minulosti. "Soukromé já" znamená uvědomování si sama sebe, vytváření myšlenek, jejichž obsahem jsou vlastní touhy, pocity, intence; a "konceptuální já" ve smyslu reflektování sebe sama v abstraktních a symbolických pojmech.

3 Závěr

Jak bylo možné vidět, vzniká široké množství pojmů pro různé typy uvědomění či vědomí. Tyto pojmy bývají homonymní, nebo se naopak sémanticky překrývají. Bylo ovšem možné je rozřadit na základě určitých shodných rysů, kterými se jednotlivé pojmy vyznačují.

Výsledkem konceptuální analýzy se stalo systematické rozdělení pojmů do takzvaných konceptuálních tříd. Vzniklé třídy jsou tedy následující: Třída 0 obsahuje koncepty, které se zaměřují na fenomenalitu vědomí. Problematičnost těchto konceptů spočívá v neshodě o tom, zda je percepční nebo i kognitivní, nebo o tom, jestli je fenomenalita vědomá, či nevědomá. Při bližším zkoumání vychází najevo, že některé koncepty se mohou shodovat s koncepty ostatních tříd, avšak liší se pouze tím, že je jim připsána fenomenální povaha. Třídy I - III potom reflektují stupeň uvědomění. Třída I zahrnuje koncepty, které reflektují neuvědomované, nebo okrajově uvědomované mentální procesy a stavy. Ty lze rozlišit podle toho, zda vycházejí ze smyslové percepce nebo jsou kognitivního rázu a jestli se jedná o vývojové stádium. Třída II je charakteristická koncepty přímého uvědomování, které odrážejí plně uvědomování si okolního prostředí. Rozlišit je lze opět podle toho, zda jsou pouze percepční, nebo i kognitivní, a do jaké míry u nich hraje roli jazyk. Ve třídě III jsou koncepty uvědomění, zaměřené na vlastní prožívané stavy nebo na uvědomění, soustředěné na sebe sama, které se nazývá metavědomím. Do I. až III. třídy jsou zařazeny i některé koncepty jáství, jakožto specifické aspekty uvědomění.

Užité pojmy samozřejmě není možno přesně zřetelně ohraničit v těchto čtyřech třídách. Škála typů vědomí či uvědomění tvoří spíše kontinuum. Představené čtyři konceptuální třídy je třeba chápat jako přibližné. Nicméně se jeví jako dostačující pomůcka pro analyzování a porovnání dalších konceptů. Podrobnými odlišnostmi mezi nimi se budu věnovat při dalším bádání.

Poděkování

Srdečně děkuji Mgr. Michalu Polákovi Ph.D. za odborné připomínky a konzultace. Dále mé díky za přínosné poznámky k textu patří kolegům Petře Borovské a Radku Čermákovi.

Literatura

- Armstrong, D. M. (1981). What is consciousness? In John Heil (ed.), *The Nature of Mind*. Cornell University Press.
- Bayne, T. (2007). Conscious states and conscious creatures: explanation in the scientific study of consciousness. *Philosophical Perspectives*, 21(1), 1–22.
- Berkovich-Ohana, A., & Glicksohn, J. (2014). The consciousness state space (CSS)-a unifying model for consciousness and self. *Frontiers in Psychology*, 5, 1–19.
- Block, N. (1995). On A Confusion About a Function of Consciousness. *Brain and Behavioral Sciences*, 18(2), 227–247.
- Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, 15(12), 567–75.
- Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, 15(8), 358–64.
- Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace and Co.
- Edelman, G. M. (2004). *Wider than the sky: The phenomenal gift of consciousness*. New Haven: Yale University Press.
- Farthing, W. G. (1992). The Concepts of Consciousness. In: *The Psychology of Consciousness*. Prentice Hall, 1 - 23.
- Gurwitsch, A. (2010). *The Collected Works of Aron Gurwitsch (1901-1973): Volume III: The Field of Consciousness: Theme, Thematic Field, and Margin*. Garcia-Gómez, J. (Eds)(Vol. 194). Springer Science & Business Media.
- Hobson, J. A. (2014). *Dream consciousness: Allan Hobson's New Approach to the Brain and Its Mind*. Tranquillo, N. (Ed). Springer.

- Hobson, J. A. (2009). REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews. Neuroscience*, 10(11), 803–813.
- James, W. (1950). The Stream of Thought. In: *The Principles of Psychology*, vol. 1, Dover, 224-290.
- Johanson, M., Revonsuo, A., Chaplin, J., & Wedlund, J. - E. (2003). Level and contents of consciousness in connection with partial epileptic seizures. *Epilepsy & Behavior*, 4(3), 279–285.
- Levine, J. (1983). “Materialism and qualia: the explanatory gap”. *Pacific Philosophical Quarterly*, 64: 354-361.
- Mangan, B. (2001). Sensation’s Ghost: The Non-Sensory “Fringe” of Consciousness. *Psyche*, 7(18).
- Marvan, T., Polák, M. (2015). *Vědomí a jeho teorie*. Vyšehrad, Plzeň, 159.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83 (October):435-50.
- Neisser, U. (1997). The roots of self-knowledge: perceiving self, it, and thou. In: J. G. Snodgrass & R. L. Thompson (Eds.), *The self across psychology: Self-recognition, self-awareness, and the self-concept*. New York: New York Academy of Sciences, 18-33.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology* 1 (1):35-59.
- Schooler, J. W. (2002). Re-representing consciousness: dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences*, 6(8), 339–344.
- Shields, Ch. (2011). On Behalf of Cognitive Qualia. In: Bayne, T., Montague, M. (eds.)(2011). *Cognitive Phenomenology*. New York: OUP.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, 26(1), 1.
- Vandekerckhove, M., & Panksepp, J. (2009). The flow of anoetic to noetic and autonotic consciousness: A vision of unknowing (anoetic) and knowing (noetic) consciousness in the remembrance of things past and imagined futures. *Consciousness and Cognition*, 18(4), 1018–1028.
- Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences*, 7(5), 214–218.
- Zelazo, P. (1996). Towards a characterization of minimal consciousness. *New Ideas in Psychology*, 14(1), 63–80.
- Zelazo, P. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences*, 8(1), 12–17.

Vedecké intuície a vedecká gramotnosť u slovenských stredoškôľákov

Vladimíra Čavojová

Katedra pedagogickej a školskej psychológie
Pedagogická fakulta UKF
Dražovská cesta 4, 94901 Nitra
vkavojova@ukf.sk

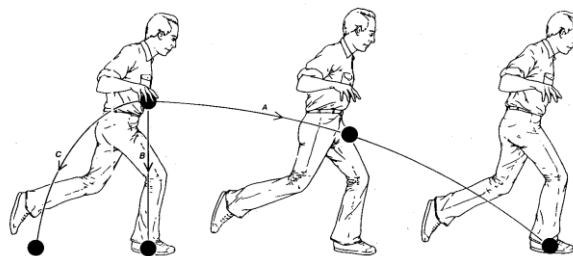
Abstrakt

Cieľom štúdie bolo replikovať zistenia Shtulmana a Valcarcela (2012), ktorí ukázali, že vedecké teórie nenahrádzajú intuitívne teórie a obohatiť ich o zistenie vzťahu s vedeckým myslením, pretože nás zaujímalo, či schopnosť porozumieť vedeckej metodológii bude mať pozitívny efekt na vedeckú gramotnosť študentov. Výskumu sa zúčastnilo 100 stredoškôľákov ($M=17,84$ rokov, $SD=0,63$), ktorí vyplnili Test vedeckého myslenia (Drummond & Fischhoff, 2015) a verifikovali 200 tvrdení v 10 vedeckých oblastiach (astronómia, evolúcia, zlomky, genetika, patogény, hmota, mechanika, fyziológia, termodynamika a vlny). Zistenia Shtulmana a Valcerala sme verifikovali, no vzťah medzi vedeckým myslením a vyššou úspešnosťou v riešení konzistentných či nekonzistentných položiek sú nepotvrdili, čo naznačuje, že vedecké poznatky sú nezávislé od schopnosti vedecky uvažovať, no tiež na nízku úroveň kriticky rozmýšľať o metódach, akými vedecké poznatky získavame.

1 Vedecké vs. intuitívne teórie sveta

Jedna za najznámejších teórií kognitívneho vývinu od J. Piageta predpokladá, že deti postupujú ako malí vedci a postupne nahrádzajú menej dokonalé teórie sveta okolo seba dokonalejšími, ktoré sú viac v zhode s vedeckým poznaním sveta. Táto teória má však mnohých kritikov (napr. Gardner, 1995) a nedávna štúdia Shtulmana a Valcerela (2012) ukázala, že intuitívne teórie nie sú vedeckými teóriami nahrádzané, ale iba potlačené, a že vysokoškolskí študenti, ktorí absolvovali kurzy prírodovedných predmetov, sú ovplyvnení týmito intuitívnymi teóriami z detstva pri časovom strese. Učítelia sa každodenne v triedach stretávajú s tým, že žiaci nie sú nepopísané tabule či prázdne nádoby, do ktorých stačí naliať poznatky, ale naopak, prichádzajú do tried s vlastným vnímaním a poznaním, ako svet okolo nich funguje a tieto intuitívne teórie často interferujú s novými poznatkami, a preto je nevyhnutné brať pri vyučovaní do úvahy (Gardner, 1995).

Napríklad McCloskey (1983) ukázal, že aj študenti fyziky majú často predstavy o mechanike, ktoré odporujú realite, ale zhodujú sa skôr s naivnou fyzikou a s vnímaním mechaniky z čias pred Newtonom. McCloskey (1983) sa napríklad pýtal svojich študentov, kam dopadne lopta, ktorú pustí človek pohybujúci sa dopredu stabilnou rýchlosťou (Obr. 1). Len 45% študentov vedelo, že lopta dopadne na mieste a (vpredu), 49% si myslelo, že lopta spadne rovno dole a pristane presne pod miestom, z ktorého bola vypustená (b) a 6% si myslelo, že spadne dozadu (c).



Obr. 1: Na ktoré miesto (a,b,c) dopadne lopta, ktorú pustí človek v pohybe? Intuitívne presvedčenia nie vždy korešpondujú s fyzikálnou realitou. Zdroj: (McCloskey, 1983)

Tieto výsledky sa však netýkajú len študentov či dospelých, ktorí možno nedávali dobrý pozor na hodinách fyziky. Napríklad Goldberg a Thompson-Schill (2009) zistili, že dokonca aj profesori biológie sú pomalší a menej presní v klasifikácii rastlín ako živých organizmov na rozdiel od zvierat ako živých organizmov, čo naznačuje, že ani mnohé roky profesionálnych skúseností nedokážu celkom vymazať intuitívne ontologické teórie, ktoré si vytvárame v detstve.

Tieto zistenia majú závažné a pesimistické dôsledky aj pre vyučovanie, v podobe akej prebieha v našich školách. Keďže školy na Slovensku stále vychádzajú prevažne z tradície jednotnej školy, vyučovanie prebieha väčšinou osvojovaním si pomerne veľkého množstva encyklopedických vedomostí a aj vyučovanie vedeckých disciplín prebieha skôr formou predávania informácií a poučiek, bez výraznejšieho pochopenia procesu, ktorým tieto vedecké poznatky vznikajú. Ak aj

profesori biológie sú po rokoch skúseností ovplyvňovaní teóriami, ktoré si vytvorili ako deti, akú šancu máme, že študenti sa budú riadiť vedeckými informáciami, s ktorými ich oboznamujeme v škole, ak tieto nestoja na hlbšom pochopení a bez neustáleho posilňovania po odchode zo školy aj tak prevážia intuitívne teórie z detstva?

Cieľom štúdie bolo preto zmapovať, ak sú na tom s vedeckou gramotnosťou slovenskí stredoškólači a replikovať tak zistenia Shtulmana a Valcarela (2012), že aj napriek pôsobeniu školy, intuitívne teórie nie sú nahradené sofistikovanejšími a správnejšími vedeckými poznatkami, ale sú len potlačené. Navyše, zaujímalo nás¹, či existuje vzťah medzi lepšou vedeckou gramotnosťou (mám viac a presnejších vedeckých poznatkov z viacerých oblastí) a schopnosťou vedecky myslieť, pretože predbežné výsledky Drummondovej a Fischoffa (2015) naznačujú, že vyššie schopnosti vedeckého myslenia vedú k lepšiemu pochopeniu vedeckých informácií.

2 Metódy

2.1 Participanti

Výskumu sa zúčastnilo 100 stredoškólačkov z 3 vybraných škôl v nitrianskom kraji. Priemerný vek študentov bol 17,84 (SD=0,63) rokov, chlapcov bolo 53 a dievčat 47. Všetci bolo testovaní počas jednej vyučovacej hodiny a za účasť dostali malú odmenu.

2.2 Materiály

Použili sme presne ten istý materiál ako Shtulman a Valcarel (2012). Participanti mali určovať pravdivosť alebo nepravdivosť 200 tvrdení o rôznych prírodných úkazoch v 10 oblastiach (astronómia, evolúcia, zlomky, genetika, patogény, hmota, mechanika, fyziológia, termodynamika a vlny). V každej oblasti sme sa pýtali na 5 konceptov, pričom štvrtina tvrdení bola pravdivá podľa intuitívnej aj vedeckej teórie v danej doméne (napr. Ľudia sú bližšie príbuzní ľudoopom než opiciam.), štvrtina tvrdení bola nepravdivá podľa intuitívnej aj vedeckej teórie (Veľryby sú príbuznejšie rastlinám ako rybám.), štvrtina tvrdení bola pravdivá podľa intuitívnej, ale nepravdivá podľa vedeckej teórie (Ľudoopy sú príbuznejšie opiciam než ľuďom.) a napokon štvrtina tvrdení bola pravdivá podľa vedeckej, ale nepravdivá podľa intuitívnej teórie (Veľryby sú príbuznejšie ľuďom než rybám.). Príklady tvrdení sú v tabuľke 1. Tvrdenia boli vyvážené z hľadiska 3 otázok: (1) v každej doméne bol rovnaký počet pravdivých aj nepravdivých tvrdení, aby sa predišlo response bias, (2)

komplexnosť tvrdení bola vyvážená v druhoch odpovedí (konzistentné vs. nekonzistentné) aj typoch odpovedí (pravdivé vs. nepravdivé), aby boli rovnako často zastúpené jednoduché aj komplexné tvrdenia, (3) priemerný počet slov v tvrdeniach bol rovnaký pre každý typ odpovedí aj doménu.

Na rozdiel od Shtulmana a Valcarela participanti odpovedali spôsobom papier-ceruzka, čiže sme nemerali čas, iba správnosť odpovedí. Správnosť bola určená podľa vedeckej teórie. Tvrdenia boli administrované participantom v randomizovanom poradí, ale po jednotlivých doménach, aby sme im uľahčili vyplňanie.

Na meranie schopnosti vedecky myslieť sme použili Test vedeckého myslenia (Drummond & Fischhoff, 2015), ktorý sa skladá z 11 modelových situácií opisujúcich nejaký vedecký koncept (double-blind random trial procedúra, kauzalita, vmedzerenné premenné, konštruktová validita, kontrolné skupiny, ekologická validita, história, maturácia, náhodné priradovanie do podmienok, response bias). Úlohou participantov je ohodnotiť, či je výrok pravdivý alebo nepravdivý. Príklad testovania konceptu kauzality:

Výskumník zistí, že v slovenských mestách s väčšími parkami žije menej ohrozených druhov zvierat. Je nasledujúci výrok pravdivý alebo nepravdivý? Tieto údaje naznačujú, že keď zväčšíme plochu parku v slovenských mestách, tak znížime počet ohrozených druhov. (T/F)

| Doména (koncept) | T1 | T2 | Tvrdenie |
|-------------------------------|----|----|---|
| Evolúcia (Spoloční predkovia) | T | T | Ľudia sú bližšie príbuzní ľudoopom než opiciam. |
| | F | F | Veľryby sú príbuznejšie rastlinám ako rybám. |
| | T | F | Ľudoopy sú príbuznejšie opiciam než ľuďom.* |
| | F | T | Veľryby sú príbuznejšie ľuďom než rybám.* |

Tab. 1: Príklady položiek z jednej domény.

Konzistentné sú tie položky, ktoré pravdivé alebo nepravdivé aj podľa intuitívnej (T1) aj podľa vedeckej (T2) teórie. Nekonzistentné sú položky, ktoré sú podľa jednej z teórií pravdivé, ale podľa druhej nepravdivé (označené hviezdičkou*).

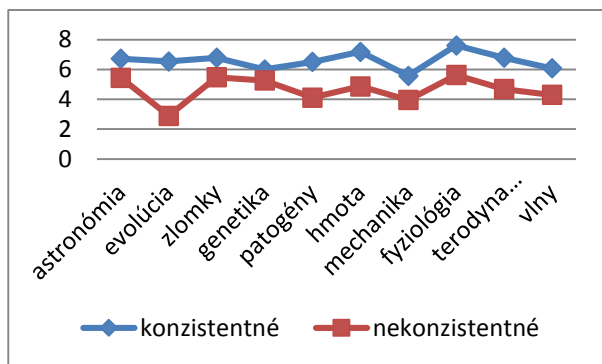
3 Výsledky

Presnosť participantov pri overovaní platnosti tvrdení, ktorých pravdivosť bola buď konzistentná alebo nekonzistentná podľa vedeckej a intuitívnej teórie, vidno na Obr. 2.

Aj výsledky t-testov pre závislý výber ukázali štatisticky významné rozdiely v prospech väčšej úspešnosti pri konzistentných tvrdeniach (Tab. 2).

¹ Tento príspevok vznikol v spolupráci s Katarínou Vargovou, ktorá zbierala dáta na svoju bakalársku prácu s názvom Vedecké intuície u stredoškólačkov.

Napokon sme počítali Pearsonovu koreláciu medzi konzistentnými a nekonzistentnými položkami v jednotlivých doménach testu vedeckých poznatkov (Shtulman & Valcarcel, 2012) ako aj celkovým skóre (súčet konzistentných a nekonzistentných položiek, ktorý odrážal celkovú mieru vedeckých poznatkov vo všetkých 10 doménach) a výsledkom v Teste vedeckého myslenia (Drummond & Fischhoff, 2015). Výsledky možno vidieť v Tabuľke 3.



Obr. 2: Priemerný počet správnych odpovedí v konzistentných a nekonzistentných položkách pre každú doménu

| dimenzia | konzistentné | | nekonzistentné | | t | p | d |
|------------|--------------|------|----------------|------|--------|-------|------|
| | M | SD | M | SD | | | |
| astronómia | 6,71 | 1,79 | 5,44 | 1,28 | 6,455 | <0,01 | 0,82 |
| evolúcia | 6,53 | 1,93 | 2,87 | 1,56 | 15,610 | <0,01 | 2,09 |
| zlomky | 6,77 | 2,69 | 5,47 | 2,41 | 7,120 | <0,01 | 0,51 |
| genetika | 5,99 | 1,96 | 5,26 | 1,81 | 3,356 | <0,01 | 0,39 |
| patogény | 6,54 | 2,36 | 4,10 | 1,91 | 11,274 | <0,01 | 1,14 |
| hmota | 7,20 | 2,51 | 4,89 | 1,81 | 9,941 | <0,01 | 1,06 |
| mechanika | 5,57 | 1,80 | 3,15 | 1,28 | 14,170 | <0,01 | 1,55 |
| fyziológia | 7,60 | 2,03 | 5,61 | 1,81 | 9,645 | <0,01 | 1,03 |

| | | | | | | | |
|------------------|------|------|------|------|--------|-------|------|
| terodyn amika | 6,79 | 1,78 | 4,67 | 1,89 | 10,767 | <0,01 | 1,15 |
| vlny | 6,10 | 2,14 | 3,71 | 1,96 | 9,862 | <0,01 | 1,16 |

Tab. 1: Porovnanie úspešnosti v konzistentných a nekonzistentných položkách vo všetkých doménach. M=priemer, SD=štandardná odchýlka, t=hodnota t-testu, p=statistická významnosť, d=Cohenovo d

| | | 1. | 2. | 3. |
|-------------------------------------|---|--------|--------|-------|
| 1. konzistentné | r | | | |
| | p | | | |
| 2. nekonzistentné | r | ,731** | | |
| | p | <0,001 | | |
| 3. test vedeckých poznatkov (spolu) | r | ,946** | ,913** | |
| | p | <0,001 | <0,001 | |
| 4. test vedeckého myslenia | r | 0,135 | 0,091 | 0,124 |
| | p | 0,18 | 0,369 | 0,218 |

Tab. 2: Korelácie medzi konzistentnými a nekonzistentnými položkami testu vedeckých poznatkov a testom vedeckého myslenia

4 Diskusia

Cieľom príspevku bolo replikovať zistenia Shtulmana a Valcarela (2012), že keď sa študenti naučia nové teórie, ktoré sú v konflikte s ich intuitívnymi teóriami, tak tieto nové teórie vedecké intuície nenahrádzajú, len ich potláčajú. Aj naše výsledky, tentokrát na vzorke stredoškôľakov, poukázali na očakávaný vzorec odpovedí -- pri tvrdeniach, kde boli v konflikte vedecké intuície s vedeckou teóriou, bola úspešnosť významne nižšia ako pri tvrdeniach, kde boli obe teórie v súlade. Silný vzťah medzi úspešnosťou riešenia konzistentných a nekonzistentných položiek poukazuje na to, že študenti, ktorí majú vo všeobecnosti viac vedeckých informácií, sú aj úspešnejší v riešení problémov, kedy sú intuitívne teórie v konflikte s vedeckými. Na druhej strane, absencia vzťahu so schopnosťou vedecky uvažovať, naznačuje, že môže ísť len o mechanicky osvojené poznatky bez hlbšieho pochopenia procesu získavania vedeckých poznatkov. Stredoškôľakov ako vzorku sme vybrali aj preto, že sme predpokladali, že sú v škole vystavení vedeckým informáciám na prírodovedných predmetoch a preto sme overili, dodatočný predpoklad, či typ školy zohráva rolu v úspešnosti odpovedí na konzistentné a nekonzistentné položky. Zistili sme, že gymnazisti sa

nelišili od študentov strednej odbornej školy ani v teste vedeckého uvažovania, ani v teste vedeckých vedomostí (s jedinou výnimkou, keď boli gymnazisti úspešnejší v riešení nekonzistentných položiek z astronómie).

5 Záver

Výsledky našej štúdie nie sú obzvlášť prekvapujúce, pretože aj v testovaní PISA sa konzistentne naši žiaci v prírodovednej gramotnosti umiestňujú na posledných miestach rebríčka (Ferencová, Stovíčková, & Galádová, 2012; Koršňáková, 2009). Alarmujúce však je, že v meraniach PISA sa ukazuje, že nielenže sa výkon našich žiakov pohybuje pod priemerom OECD, ale že sa v poslednom cykle merania aj významne zhoršil.

V našom výskume však nešlo len o zmeranie úrovne vedeckých poznatkov stredoškôľakov, ale aj o preskúmanie vzťahu medzi vedeckými poznatkami a vedeckým uvažovaním, pretože podľa Drummondovej a Fischhoffa (2015) vedecké uvažovanie napomáha chápanou a akceptácii vedeckých informácií. Ukázalo sa, že žiaci majú nedostatočnú úroveň vedeckého uvažovania (v priemere dosahovali len 6 bodov z 11, čo môže nasvedčovať, že mnohé odpovede len tipovali). Absencia vzťahu medzi lepším vedeckým uvažovaním a vyššou úspešnosťou pri verifikácii vedeckých tvrdení z testu Shtulmana a Valcarela naznačuje, že aj úspešnejší žiaci v skutočnosti majú len viac vedomostí v danej doméne, bez výraznejšieho pochopenia procesu, ktorý viedol k získaniu týchto poznatkov.

Na záver by sme mohli zhrnúť, že vyučovanie (nielen) prírodovedných predmetov by malo prebiehať skôr spôsobom oboznamujúcim žiakov s metódami vedeckého bádania ako memorovaním poučiek z učebníc. Pochopenie toho, ako poznanie vzniká, vedie ku kritickjšiemu zhodnoteniu vedeckých výsledkov v súčasnosti a menšiemu podliehaniu rôznym nepodloženým pseudovedeckým tvrdeniam. V dnešnej dobe, keď sú informácie na dosah ruky (resp. vzdialené len na pár klikov myšou) je nevyhnutné učiť žiakov kriticky myslieť. Stávame sa čoraz závislejší na výsledkoch vedeckého bádania a technologického pokroku, preto je potrebné aspoň základné pochopenie princípov vedeckého myslenia. Veríme, že pochopenie procesu, akým vznikajú vedecké poznatky, by malo viesť k väčšej dôvere v štandardnú vedu a väčšiu skepsu voči rôznym zázračným kúram, ako napríklad použitiu magnetu na "odblokovanie" emócií, pitie bielidla na liečbu rakoviny či sódom bikarbónou proti "prekysleniu" organizmu.

PodĎakovanie

Tento príspevok vznikol za podpory grantovej agentúry VEGA v rámci grantovej úlohy 1/0028/15: Výchova k racionalite: Kognitívne schopnosti, dispozície myslenia a výsledky v reálnom živote.

Literatúra

Drummond, C., & Fischhoff, B. (2015). Predicting the reception of scientific evidence: The role of Emotion and scientific reasoning ability. In SPUDM 2015. Budapest.

Ferencová, J., Stovíčková, J., & Galádová, J. (2012). Národná správa PISA 2012. Bratislava.

Gardner, H. (1995). *The Unschooled Mind. How children think and how schools should teach.* New York: Basic Books, Inc., Publishers.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge: Research Article. *Psychological Science*, 20(4), 480–487. doi:10.1111/j.1467-9280.2009.02320.x

Koršňáková, P. (2009). Pisa 2009. Organization for Economic Co-operation and Development. Bratislava. Retrieved from <http://pisa2009.acer.edu.au/downloads.php>

McCloskey, M. (1983). Intuitive physics. *Scientific American*, 248(4), 122–130. doi:10.1038/scientificamerican0483-122

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124(2), 209–215. doi:10.1016/j.cognition.2012.04.005

Čo z čoho a kto z koho?

Efekt číselného vyjadrenia a odbornosť pri posudzovaní rizika v oblastiach zdravia a financií

Katarína Dudeková, Lenka Kostovičová

Ústav experimentálnej psychológie CSPV SAV
Dúbravská cesta 9, 84101 Bratislava
lenka.kostovicova@savba.sk, katarina.dudekova@savba.sk

Abstrakt

Ľudia sa pri posudzovaní rizík dopúšťajú kognitívnych odchýlok. Jednou z nich je vnímanie matematicky ekvivalentných informácií, prezentovaných v rôznych formátoch, odlišne. Platí to i pre finančných profesionálov? V experimentálnej štúdií ($N = 99$) s 3 (formáty) \times 2 (domény) dizajnom sme overovali rozdiely v subjektívnej miere pravdepodobnosti, závažnosti a obáv vzhľadom na číselné vyjadrenie informácie. V oblasti zdravia sme identifikovali fenomén „1 z X“, kedy tento formát viedol k najvyššej úrovni vnímaného rizika. Efekt nastal vďaka tendencii predstaviť si pod osobou v zadaní samého seba a nie iných. Vo finančnej doméne formát vnímanie rizika neovplyvnil. Venujeme sa aplikáciám zistení, najmä v oblasti komunikácie štatistických informácií.

1 Posudzovanie rizika

Adekvátne posúdenie úrovne rizika je základom informovaných rozhodnutí, ktoré majú častokrát nielen individuálny, ale i spoločenský dosah. Prostredníctvom miery vnímaných rizík dokážeme predikovať reálne správanie ľudí (napr. Falk a spol., 2010; Weinstein a spol., 2007). Preto je dôležité zamerať sa na faktory, ktoré zohrávajú rolu v posudzovaní pravdepodobnosti a rizika. Jedným z nich je formulácia informácií.

1.1 Vplyv formátu na posúdenie rizika

Nielen pri slovných vyjadreniach, ale i pri číslach platí, že tá istá informácia, prezentovaná odlišným spôsobom, je chápaná ináč. Najlepšie dokumentované efekty číselného formátu na vnímanie rizika sú zanedbanie menovateľa a skupinová difúzia.

Efekt zanedbania menovateľa spočíva v prílišnom zameraní sa na čitateľa, nedostatočne zohľadňujúc základný výskyt, respektíve veľkosť referenčnej skupiny (Denes-Raj a spol., 1995). V dôsledku tejto tendencie ľudia volia nádobu s 10 výhernými

gulôčkami zo 100 radšej než nádobu s 1 výhernou gulôčkou z 10, hoci ide o rovnocenné možnosti. Pred druhou nádobou však uprednostnia často i možnosť 7 z 10, dopúšťajúc sa tak iracionálnej kognitívnej odchýlky. Fenomén sa v literatúre nazýva i efekt pomerového skreslenia.

Protichodné zistenia priniesol výskum o efekte skupinovej difúzie. Je založený na predpoklade "bezpečnosť je v číslach". Teda, so stúpajúcim počtom ľudí, ktorým hrozí určité riziko, klesá pravdepodobnosť, že hrozba postihne konkrétneho človeka (Yamaguchi, 1998). V dôsledku tohto sklonu ľudia vnímajú riziko ako nižšie ak sa týka skupiny 100 ľudí, ako v prípade potenciálneho ohrozenia 10 ľudí.

V posledných rokoch niekoľko výskumov preukázalo existenciu tretieho, komplementárneho efektu, ktorý sa s predošlými dvoma čiastočne (skupinová difúzia) alebo úplne (zanedbanie menovateľa) vylučuje.

1.2 Efekt "1 z X"

Pighin a spol. (2011) v oblasti posudzovania zdravotných rizík zistili, že si najvyššiu pravdepodobnosť ľudia spájajú s formátom 1 z X. Efekt overili za použitia rôznych číselných pomerov, diagnóz, i vzoriek participantov. V rámci nadväzujúcej štúdie so šiestimi experimentmi Sirota a spol. (2014) prostredníctvom meta-analýzy potvrdili existenciu tohto efektu, avšak v obmedzenej miere, pričom časť ich experimentov nevedla k významným výsledkom.

1.3 Posudzovanie rizika expertmi

Skreslené vnímanie rizík sa vyskytuje bez ohľadu na mieru vzdelanosti či vedeckej gramotnosti (Kahan a spol., 2012). V posudzovaní pravdepodobnosti bežne zlyhávajú i odborníci v oblasti medicíny či práva (Gigerenzer a spol., 1995; Goodman-Delahunty a spol., 2010). Čo sa týka finančných profesionálov, tí bývajú spravidla vo svojich odhadoch nadmerne optimistickí a

miskalibrovaní (Glaser a kol., 2005; Grežo, 2015). Výsledky nášho nedávneho výskumu však naznačujú, že expertiza v doméne by mohla znižovať efekt niektorých kognitívnych skreslení v usudzovaní (Dudeková & Kostovičová, 2015). Rovnako sme zistili, že znalosť témy znižuje neadekvátnu mieru vnímaných rizík (Kostovičová, 2016). Preto sme sa rozhodli overiť naše predpoklady na vzorke finančných profesionálov a zistiť, či podliehajú efektu formulácie informácií v doménach, z ktorých jedna sa týka ich pracovnej náplne a v druhej by sa dostatočne vyznať nemali.

1.4 Výskumný cieľ a hypotézy

V našej experimentálnej štúdií na báze upravenej repliky výskumu Sirotu a spol. (2014) sme sa rozhodli testovať predpoklady v súvislosti s každým z troch efektov, v dvoch doménach. Zároveň je našim cieľom overiť vysvetlenie efektu "1 z X", navrhnuté autormi Pighin a spol. (2011). Zaujímá nás vplyv formátu na posúdenie pravdepodobnosti, závažnosti a obáv. V rámci vysvetlenia sa zameriame i na predstavu samých seba verzus iných u jednotlivých formátov.

Pre účel formulácie hypotéz si zdefinujeme tri formáty vyjadrenia identickej informácie: 1 z X, n z nX a m z mX, pričom $1 < n < m$.

Hypotézu zanedbania menovateľa môžeme vyjadriť ako $1 \text{ z } X < n \text{ z } nX < m \text{ z } mX$. Teda vnímaná miera pravdepodobnosti, závažnosti a obáv so stúpajúcou hodnotou čitateľa narastá.

Predpoklad skupinovej difúzie môžeme formulovať ako $1 \text{ z } X > n \text{ z } nX > m \text{ z } mX$. Teda vnímaná miera pravdepodobnosti, závažnosti a obáv so stúpajúcou hodnotou menovateľa klesá.

Podľa princípov efektu "1 z X" by malo platiť $1 \text{ z } X > n \text{ z } nX = m \text{ z } mX$. Teda vnímaná miera pravdepodobnosti, závažnosti a obáv je najvyššia u formátu 1 z X a u ďalších dvoch formátov sa nelíši.

V rámci možného vysvetlenia efektu "1 z X" sa pýtame, či k zmene vo vnímanej miere pravdepodobnosti, závažnosti a obáv dochádza v dôsledku rozličnej miery predstáv samého seba verzus iných ľudí vplyvom rôznych formátov. Teda je predstava samého seba mediátorom efektu formátu na vnímanie rizika a iných aspektov?

2 Metódy

2.1 Participanti

Na výskume participovalo 99 profesionálov z oblasti finančnictva vo veku 21 až 62 rokov ($M = 37,5$; $SD = 9,4$). Muži tvorili 54,5% výskumného súboru a ženy

45,5%. Dĺžka praxe sa pohybovala v rozmedzí 0,5 - 34 rokov ($M = 9,3$; $SD = 8,1$). Medzi profesiami participantov sa najčastejšie objavovali nasledovné: finančný konzultant, finančný poradca, finančný sprostredkovateľ, finančný agent, finančný analytik, účtovník a ekonóm.

2.2 Dizajn a operacionalizácia

V rámci medzi-subjektového experimentálneho dizajnu boli participanti náhodne pridelení do jednej z troch skupín: formát 1 z X, formát n z nX a formát m z mX, pričom $1 < n < m$. Konkrétne išlo o 1 z 12 ($n = 33$), 5 zo 60 ($n = 35$) a 10 zo 120 ($n = 31$). Svoje vyjadrenia robili v dvoch oblastiach (financie a zdravie).

2.3 Materiály a procedúra

V oblasti financií si participanti prečítali stručnú informáciu o akciovom fonde v rámci sporenia v II. dôchodkovom pilieri a v oblasti zdravia popis diagnózy nespavosť - insomnia:

i) Financie: *V rámci II. piliera má sporiteľ možnosť investovať rozložením svojich úspor do dlhopisového a akciového fondu. Akciový fond predstavuje vyššie riziko ako dlhopisový. Jeho výnos je vysoký, nestabilný a môže výraznejšie kolísť.*

ii) Zdravie: *Nespavosť (odborne insomnia) je stav, pri ktorom človek zle zaspáva, často sa budí a má plytký spánok. Príčiny nespavosti sú rôzne: od nezávažných - napr. zmena prostredia či vyčerpanie, až po závažné - napr. zlyhanie pľúc alebo chronická bolesť.*

Nasledoval výrok, ktorý obsahoval štatistickú informáciu v jednom z troch formátov. Napríklad pri formáte „1 z X“ zneli výroky nasledovne:

i) Financie: *Riziko, že dospelý Slovák / dospelá Slovenka, sporiaci/-a v rámci II. piliera v akciovom fonde, utrpí v určitej fáze investovania výrazný pokles hodnoty úspor, je 1 z 12.*

ii) Zdravie: *Riziko, že dospelý Slovák / dospelá Slovenka bude v určitej fáze života trpieť nespavosťou, je 1 z 12.*

Participanti sa na základe prečítaných informácií vyjadrovali k výrokom o vnímanom riziku, závažnosti a obavách. Pravdepodobnosť, že dospelý Slovák (dospelá Slovenka), sporiaci/-a v rámci II. piliera v akciovom fonde, utrpí v určitej fáze investovania výrazný pokles hodnoty úspor / bude v určitej fáze života trpieť nespavosťou / hodnotili experti na Likertovej škále 1 (veľmi nízka) až 10 (veľmi vysoká). K závažnosti týchto situácií sa vyjadrovali na škále s koncovými bodmi 1 (úplne nezávažné) až 10 (úplne závažné) a obavy v súvislosti s poklesom úspor a predstretou diagnózou

vyjadrovali pomocou škály 1 (žiadne obavy) až 10 (veľké obavy).

Následne sme sa finančníkov pýtali, koho si predstavili na mieste osoby, ktorej hrozí strata / diagnóza. Svoj úsudok vyjadrovali na škále 1 (predstavil/-a som si seba) až 10 (predstavil/-a som si iných). Nakoniec sa účastníci vyjadrovali, či investujú v rámci svojho dôchodkového sporenia v II. pilieri do akciových fondov a či v minulosti oni alebo niekto blízky z ich rodiny trpeli diagnózou nespavosti.

Poradie administrácie scenárov z oblastí financie a zdravie bolo náhodné. Po vyplnení všetkých, vrátane socio-demografických, údajov, mali účastníci možnosť v osobitnom dokumente uviesť svoju emailovú adresu, kde im bola neskôr doručená správa z výskumu.

3 Výsledky

Úroveň vnímaného rizika, závažnosti situácie, vnímaných obáv a tiež miera predstavy iných ľudí alebo samého seba ako človeka, ktorému hrozí strata / diagnóza spĺňali podmienky normality distribúcie, preto sme pracovali s hodnotami priemerov, štandardných odchýlok a parametrickými testami.

3.1 Porovnanie domén

Hodnoty priemerov (a štandardných odchýlok) vnímaného rizika (R), závažnosti (Z), obáv (O) a predstavy seba (P) pri oboch oblastiach spolu s výsledkami štatistického testovania uvádzame v Tabuľke 1. Expertná doména - akciový fond - je označená ako "FOND" a neexpertná doména - nespavosť - má označenie "NESP".

| | FOND | NESP | |
|----------|----------------|----------------|---|
| | <i>M (SD)</i> | <i>M (SD)</i> | výsledok testu |
| R | 5,83 (2,27) | 4,91 (2,07) | $t(98) = 4,28$ $p < 0,001$ $d = -0,42$ |
| Z | 4,83 (2,29) | 6,24 (2,14) | $t(98) = -5,09$ $p < 0,001$ $d = 0,64$ |
| O | 4,25 (2,46) | 5,25 (2,43) | $t(98) = -3,75$ $p < 0,001$ $d = 0,41$ |
| P | 5,8 (3,10) | 5,8 (3,01) | $t(98) < 0,000$ $p = 1,000$ $d < 0,01$ |

Tab. 1: Výsledky porovnania domén

Finančníci posúdili riziko straty v rámci investovania v akciovom fonde (expertná doména) ako významne väčšie, pričom jej pripisovali výrazne menšiu závažnosť a obavy ako pri posudzovaní nespavosti.

Možno tvrdiť, že výsledok odzrkadľuje rozhradenosť finančných profesionálov v oblasti investovania v akciových fondoch.

3.2 Kontrolné premenné: rod, poradie, osobná relevancia, vek, dĺžka praxe.

Pri porovnaní na základe kontrolných premenných sme objavili tri skupiny významných výsledkov - podľa rodu, poradia administrovaných materiálov a na základe osobnej relevancie. Poradie materiálov, dĺžka praxe a vek nesúviseli ani s jednou z našich kľúčových závislých premenných v doménach zdravia a financií. Pri skupinách porovnaní sme aplikovali Bonferroniho úpravu hladín významnosti.

V súvislosti s rodom sme identifikovali výrazné rozdiely pri posudzovaní obáv zo straty v rámci investovania v akciovom fonde II. dôchodkového piliera. Ženy sa straty hodnoty akcií obávali viac ($M = 4,98$; $SD = 2,50$) ako muži ($M = 3,65$; $SD = 2,28$), $t(97) = 2,76$; $p = 0,007$; $d = 0,67$. Výsledok je v súlade so zisteniami o rodových rozdieloch v rizikových postojoch (Dudeková, 2014).

V akciovom fonde II. piliera investuje 56,6% našich účastníkov. V porovnaní s neinvestujúcimi táto časť výskumného súboru vnímala nižšiu mieru závažnosti poklesu hodnoty akcií ($M = 5,53$; $SD = 2,28$ verzus $M = 4,29$; $SD = 2,16$), $t(97) = 2,78$; $p = 0,007$; $d = 0,56$. Rozdiel pripisujeme vyššej znalosti vývoja cien akcií vďaka osobnej relevancii.

Nespavosť sa týka 29,3% našich účastníkov alebo ich rodinných príslušníkov. V porovnaní s účastníkmi, pre ktorých scenár nie je osobne relevantný, vnímala táto časť výskumného súboru vyššiu mieru obáv v súvislosti s rizikom nespavosti a tiež mali vyššiu tendenciu predstaviť si seba ako človeka, ktorému hrozí diagnóza, obavy ($M = 4,69$; $SD = 2,33$ verzus $M = 6,62$; $SD = 2,16$), $t(97) = -3,84$; $p < 0,001$; $d = -0,86$; predstavy ($M = 5,19$; $SD = 3,03$ verzus $M = 7,28$; $SD = 2,42$), $t(97) = -3,30$; $p = 0,001$; $d = -0,76$). Výsledky sú konzistentné s evidenciou o rodinnej anamnéze ako kľúčovom faktore v rámci vnímania zdravotných rizík (Yoon a spol., 2002).

3.3 Efekt formátu

Hodnoty priemerov jednotlivých premenných (mimo predstavy seba verzus iných) naznačovali predpoklad skupinovej difúzie, kedy vnímaná miera so stúpajúcou hodnotou menovateľa klesala (riziko: 1 z 12 - $M = 6,39$; $SD = 2,61$; 5 zo 60 - $M = 5,60$; $SD = 2,31$; 10 zo 120 - $M = 5,48$; $SD = 1,73$; závažnosť: 1 z 12 - $M = 5,12$; $SD = 2,50$; 5 zo 60 - $M = 4,86$; $SD = 2,25$; 10 zo 120 - $M = 4,48$, $SD = 2,13$; obavy: 1 z 12 - $M = 4,70$; $SD = 2,60$; 5 zo 60 - $M = 4,23$; $SD = 2,40$; 10 zo 120 -

$M = 3,81$; $SD = 2,37$). Pri predstave seba boli priemerné hodnoty nasledovné: 1 z 12 - $M = 5,52$; $SD = 3,18$; 5 zo 60 - $M = 6,29$; $SD = 3,08$; 10 zo 120 - $M = 5,55$; $SD = 3,05$.

Ako vidíme v rámci výsledkov uvedených v nasledujúcej Tabuľke 2, formáty 1 z 12, 5 zo 60 a 10 zo 120 nevedli k výrazne odlišnej miere vnímanej pravdepodobnosti, závažnosti ani obáv zo straty v rámci investovania v akciovom fonde. U vnímanej pravdepodobnosti mierna hodnota vecnej významnosti ($d = 0,36$) ukazovala v prospech efektu „1 z X“. Výsledok v súvislosti s najnižšími obavami u formátu 10 zo 120 bol taktiež mierne vecne významný ($d = 0,27$). Celkovo sa však ani jeden z efektov – „1 z X“, zanedbania menovateľa, veľkosti referenčnej skupiny v oblasti financií nepotvrdil.

| | <i>Diff</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--|-------------|----------|----------|----------|
| Hypotéza 1a: 1 z 12 > (5 z 60, 10 zo 120) | | | | |
| Hypotéza 1b: 1 z 12 < (5 zo 60, 10 zo 120) | | | | |
| Riziko | 0,85 | 1,63 | 0,108 | 0,36 |
| Závažnosť | 0,44 | 0,90 | 0,371 | 0,19 |
| Obavy | 0,67 | 1,27 | 0,206 | 0,27 |
| Predstavy seba | -0,42 | -0,64 | 0,523 | 0,13 |
| Hypotéza 2a: 5 z 60 = 10 zo 120 | | | | |
| Hypotéza 2b: 5 z 60 > 10 zo 120 | | | | |
| Hypotéza 2c: 5 z 60 < 10 zo 120 | | | | |
| Riziko | 0,12 | 0,23 | 0,817 | 0,06 |
| Závažnosť | 0,37 | 0,69 | 0,493 | 0,17 |
| Obavy | 0,42 | 0,72 | 0,476 | 0,18 |
| Predstavy seba | 0,74 | 0,97 | 0,334 | 0,24 |

Tab. 2: Výsledky efektu formátu v doméne financií

Efekt formátu sa prejavil u diagnózy nespavosti. Už hodnoty priemerov naznačovali, že účastníci pri formáte 1 z 12 oproti ostatným formátom uvádzali vyššiu mieru pravdepodobnosti, závažnosti, obáv a predstáv seba. (pravdepodobnosť: 1 z 12 - $M = 5,55$; $SD = 2,14$; 5 zo 60 - $M = 4,66$; $SD = 2,27$; 10 zo 120 - $M = 4,52$, $SD = 1,61$; závažnosť: 1 z 12 - $M = 6,82$; $SD = 2,10$; 5 zo 60 - $M = 6,20$; $SD = 2,20$; 10 zo 120: $M = 5,68$, $SD = 2,04$; obavy: 1 z 12 - $M = 6,03$; $SD = 2,60$; 5 zo 60 - $M = 5,00$; $SD = 2,38$; 10 zo 120 - $M = 4,71$; $SD = 2,16$; predstava seba: 1 z 12 - $M = 6,67$; $SD = 2,72$; 5 zo 60 - $M = 5,89$; $SD = 3,23$; 10 zo 120: $M = 4,68$; $SD = 2,75$).

Ako zobrazujeme v Tabuľke 3, použité štatistické formáty 1 z 12, 5 zo 60 a 10 zo 120 viedli k významne odlišnej miere pri všetkých vnímaných premenných v prospech efektu „1 z X“.

| | <i>Diff</i> | <i>t</i> | <i>p</i> | <i>d</i> |
|--|-------------|----------|----------|----------|
| Hypotéza 1a: 1 z 12 > (5 z 60, 10 zo 120) | | | | |
| Hypotéza 1b: 1 z 12 < (5 zo 60, 10 zo 120) | | | | |
| Riziko | 0,96 | 2,20 | 0,03 | 0,47 |
| Závažnosť | 0,86 | 1,92 | 0,058 | 0,41 |
| Obavy | 1,17 | 2,30 | 0,024 | 0,48 |
| Predstavy seba | 1,44 | 2,29 | 0,024 | 0,50 |
| Hypotéza 2a: 5 z 60 = 10 zo 120 | | | | |
| Hypotéza 2b: 5 z 60 > 10 zo 120 | | | | |
| Hypotéza 2c: 5 z 60 < 10 zo 120 | | | | |
| Riziko | 0,14 | 0,29 | 0,77 | 0,07 |
| Závažnosť | 0,52 | 1,00 | 0,322 | 0,25 |
| Obavy | 0,29 | 0,52 | 0,607 | 0,13 |
| Predstavy seba | 1,21 | 1,62 | 0,109 | 0,40 |

Tab. 3: Výsledky efektu formátu v doméne zdravia

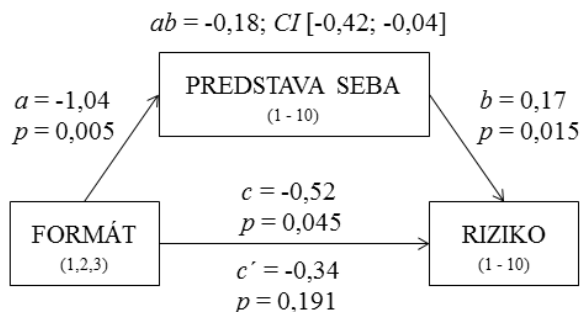
3.4 Možné vysvetlenie efektu „1 z X“

Na overenie potenciálneho vysvetlenia efektu "1 z X" na vnímanú mieru pravdepodobnosti, závažnosti a obáv pri diagnóze nespavosti prostredníctvom rozličnej miery predstáv samého seba sme použili SPSS makro PROCESS (Hayes, 2013).

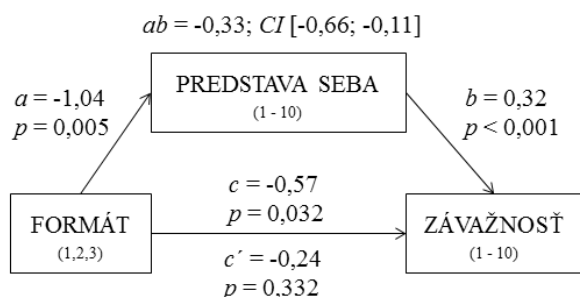
Zvyšovaním menovateľa v jednotlivých ponúkaných formátoch sa znižovala predstava seba (negatívna signifikantná cesta a, Obrázok 1, 2 a 3). Ďalej bola identifikovaná signifikantná pozitívna cesta predstavy seba ku vnímanému riziku (cesta b, Obrázok 1), ku vnímanej závažnosti (cesta b, Obrázok 2) i ku vnímaným obavám (cesta b, Obrázok 3).

Priamy efekt formátu na vnímanie rizika, závažnosti a obáv po kontrolovaní miery predstavy seba samého neostal signifikantný ani pri jednej z troch analýz (cesta c', Obrázok 1, 2, 3). Nepriame efekty boli signifikantné pri všetkých analýzach; vnímané riziko ($-0,18$; $CI [-0,42; -0,04]$), závažnosť ($-0,33$; $CI [-0,66; -0,11]$); obavy ($-0,30$; $CI [-0,67; -0,08]$). Potvrdila sa tak úplná mediácia efektu formátu na vnímanú pravdepodobnosť, závažnosť situácie ako aj na obavy prostredníctvom miery predstavy seba samého.

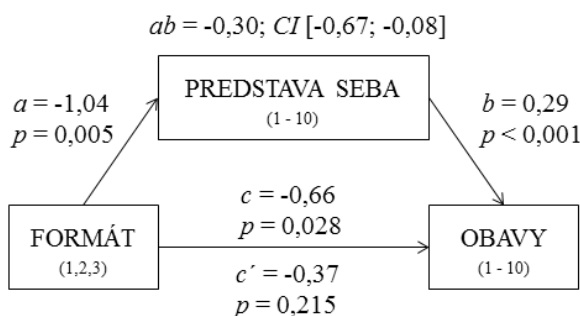
Konštatujeme, že vysvetlením efektu "1 z X" na posúdenie pravdepodobnosti, závažnosti a obáv v súvislosti s nespavosťou je zvýšená miera predstáv seba ako človeka, ktorému hrozí diagnóza.



Obr. 1: Mediačná analýza - predstava seba ako mediátor efektu formátu na vnímané riziko



Obr. 2: Mediačná analýza - predstava seba ako mediátor efektu formátu na vnímanú závažnosť



Obr. 3: Mediačná analýza - predstava seba ako mediátor efektu formátu na vnímané obavy

4 Diskusia

Predpokladom optimálnych rozhodnutí o zdraví či financiách je schopnosť ohodnotiť príslušné riziká a prínosy, ktoré pre nás predstavujú rôzne alternatívy a ich potenciálne dôsledky. V našej práci sme sa zamerali na overenie efektu formátu na vnímanie rizika, závažnosti a obáv na vzorke finančných profesionálov. Zaujímalo nás, či expertíza v jednej z domén bude viesť k redukcii efektu formulácie číselných informácií na pravdepodobnostné úsudky.

Zistili sme, že miera posúdeného rizika, závažnosti a obáv v doméne financií (t.j. investovanie v akciovom fonde II. piliera) nezávisela od číselného formátu. Výsledok podporuje naše závery o tom, že informovanosť a expertíza môžu znižovať mieru výskytu kognitívnych skreslení v usudzovaní (Dudeková & Kostovičová, 2015; Kostovičová, 2016).

V oblasti zdravia (t.j. diagnóza nespavosti) sme v súlade s replikovanou štúdiou (Sirota a spol., 2014) potvrdili efekt "1 z X". Formát s číslom "1" viedol k najvyššej miere vnímanej pravdepodobnosti, závažnosti a obáv z predmetnej diagnózy a zvyšné dva formáty sa v jednotlivých mierach nelíšili. Navyše, podarilo sa nám podporiť i vysvetlenie navrhnuté autormi Pighin a spol. (2011). Formát 1 z X spôsobil, že ľudia vnímali diagnózu ako pravdepodobnejšiu, závažnejšiu a väčšie obavy evokujúcu, pretože si pod osobou v zadaní mali tendenciu predstaviť seba a nie iných.

Náš výskum je teda nielen odpoveďou na alarmujúci nedostatok replikácií v experimentálnom psychologickom výskume (Francis, 2012), ale aj na výzvu ohľadne hľadania vysvetlení, nielen výskytu (a absencie) skúmaných vplyvov (Hayes, 2013). Navyše, okrem vysvetlenia efektu "1 z X" sa nám podarilo identifikovať i hranice jeho platnosti, respektíve jeho doménovú špecifickosť vzhľadom na oblasť expertízy usudzovateľov. Podnetným smerovaním výskumu by mohlo byť overenie našich zistení na vzorke expertov v oblasti medicíny.

Aplikácie našich výsledkov, ako i zistení výskumov o posudzovaní rizika a pravdepodobnosti vo všeobecnosti, sa týkajú najmä oblasti komunikovania informácií verejnosti a špecifickým cieľovým skupinám. I finanční profesionáli podľahli efektu "1 z X" v doméne, ktorá im nie je tak blízka ako oblasť ich profesie. Preto je dôležité zohľadňovať efekt formulácie na posudzovanie a rozhodovanie, predovšetkým ak ide o voľby so závažnými dôsledkami (Bačová, 2013). Závery výskumov o vplyve formulácie na posudzovanie pravdepodobnosti, závažnosti a obáv by sa mohli stať súčasťou manuálov a školení pre finančných agentov a sprostredkovateľov, ako aj pre lekárov, ktorí by ich využili pri komunikácii s klientmi a pacientmi, zohľadňujúc primeranú úroveň vnímaných rizík.

PodĎakovanie

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

- Bačová, V. (2013). Rozhodnutia so závažnými dôsledkami. V zborníku *Otázky rozhodovania: teória, empiria, život; 31. Psychologické dni*, str. 14–18.
- Denes-Raj, V., Epstein, S. a Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*. 21(10): 1083–1092
- Dudeková, K. (2014). *Psychologické aspekty parametrov rozhodovania v prospektovej teórii*. Bratislava: Univerzita Komenského.
- Dudeková, K. a Kostovičová, L. (2015). Oscary, HDP a CRT: Efekt ukotvenia u finančných profesionálov v kontexte doménovej špecifickosti a kognitívnej reflexie. V zborníku *Kognícia a umelý život XV*, str. 50–56.
- Falk, E. B., Berkman, E. T., Mann, T., Harrison, B. a Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *The Journal of Neuroscience*. 30(25): 8421–8424
- Francis, G (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*. 19(6): 975–991
- Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counselling for low-risk clients. *AIDS Care*. 10(2): 197–211
- Glaser, M., Langer, T. a Weber, M. (2005). *Overconfidence of professionals and lay men: Individual differences within and between tasks?* University of Mannheim.
- Goodman-Delahunty, J., Granhag, P. A., Hartwig, M. a Loftus, E. F. (2010). Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy, and Law*. 16(2): 133–157
- Grežo, M. (2015). Efekt nadmernej sebadôvery v usudzovaní a rozhodovaní finančných profesionálov. V zborníku *Cesty k rozhodovaniu: Šesť dekád psychologického výskumu*, str. 66–75.
- Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D. a Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*. 2: 732–735
- Kostovičová, L. (2016). *Testing the Cultural cognition theory of risk perception*. Bratislava: Univerzita Komenského.
- Pighin, S., Savadori, L., Barilli, E., Cremonesi, L., Ferrari, M. a Bonnefon, J. F. (2011). The 1-in-X effect on the subjective assessment of medical probabilities. *Medical Decision Making*. 31(5): 721–729
- Sirota, M., Juanchich, M., Kostopoulou, O. a Hanák, R. (2014). Decisive evidence on a smaller-than-you-think phenomenon: Revisiting the “1-in-X” effect on subjective medical probabilities. *Medical Decision Making*. 34(4): 419–429
- Weinstein, N. D., Kwitel, A., McCaul, K. D., Magnan, R. E., Gerrard, M. a Gibbons, F. X. (2007). Risk perceptions: assessment and relationship to influenza vaccination. *Health Psychology*. 26(2): 146–151
- Yamaguchi, S. (1998). Biased risk perceptions among Japanese: Illusion of interdependence among risk companions. *Asian Journal of Social Psychology*. 1(2): 117–131
- Yoon, P. W., Scheuner, M. T., Peterson-Oehlke, K. L., Gwinn, M., Faucett, A. a Khoury, M. J. (2002). Can family history be used as a tool for public health and preventive medicine? *Genetics in Medicine*. 4(4): 304–310

On abstraction: psychological, neural, and computational perspectives

Igor Farkaš

Centre for Cognitive Science
Faculty of Mathematics, Physics and Informatics, Comenius University
Mlynská dolina, 84248 Bratislava
farkas@fmph.uniba.sk

Abstract

Abstraction is a core concept in cognitive science, representing a challenge for all theories of cognition. Conceptualization of abstraction is also complicated by the fact that it is an entity with several potential meanings and involved mechanisms. Abstraction occupies the agenda of many disciplines, including psychology, linguistics, artificial intelligence, and more recently, neuroscience. In this paper, we attempt to shed light on this topic, by summarizing evidence accumulated in these disciplines.

1 Introduction

Abstraction is one of the core concepts in cognitive science. It is an important building mechanism in the cognition of intelligent beings, especially humans, helping them to efficiently cope with the complex and dynamic world. However, abstraction is not only interesting from the perspective of cognition; it appears in the agenda of many disciplines: psychology, computer science, artificial intelligence, mathematics, linguistics, music, philosophy, art, and more recently, also neuroscience. Despite evident progress in research, the concept of abstraction is not still completely understood in empirical sciences, nor sufficiently well grasped in computational models. One reason for this is that abstraction is not a single homogeneous entity, but rather offers several interpretations and underlying mechanisms.

Abstraction can be primarily seen as a conceptual process by which general rules and concepts are derived from the usage and classification of specific examples. Conceptual abstractions may be formed by filtering the information content of a concept or an observable phenomenon, selecting only those aspects which are relevant for a particular purpose. In this paper, we do not attempt to provide a complete survey of this complex topic. Instead, we try to summarize evidence (not exhaustively) coming from the disciplines crucial for cognitive science (psychology, linguistics, neuroscience and computer science).

Abstraction typically refers to representing abstract concepts that can be mostly expressed by words in natural language(s). It can also refer to symbols (in

Peirce's semiotic view) that have meanings in certain worlds (e.g. mathematics) or cultures. Linguistic symbols (words) can refer not only to objects (expressed by nouns), but also actions (verbs), or properties (adjectives). Within cognitive science, abstraction is differently viewed in classical theories and in embodied theories of cognition.

1.1 Theories of cognition

According to the *classical view of cognition*, the mind is considered a symbol system and cognition relates to processes of symbol manipulation (Fodor, 1998). Cognition and action/perception are separate and independent systems that work according to different principles (Barsalou, 1999). Hence, conceptual representations are non perceptual and unrelated to the body. In this framework, concepts are generated by combining and manipulating abstract, arbitrary and amodal symbols for which their internal structures are unrelated to the perceptual states and actions that produced them (Fodor, 1998).

The problem of classical approach to assign meanings to symbols, articulated by the Chinese room argument (Searle, 1980), and subsequently by the symbol grounding problem (Harnad, 1990), triggered the development of alternative theories of *grounded cognition* (see the overview, e.g. in Barsalou 1999). In these views, conceptual knowledge is represented with (multi)modal symbols related to the perceptual states that produce them. The concept of grounding embraces the other two crucial concepts – embodiment and situatedness. *Embodiment* provides the pathways towards learning concepts via sensorimotor interaction with the environment, enabled by sensory-motor features (perceptual and proprioceptive) originating in agent's sensors and actuators. *Situatedness* provides rich repertoire of features from the environment about which particular knowledge is being acquired (situated learning).

Within the grounded cognition paradigm, cognitive linguistic theories deny the presence of a separate and autonomous language module in the brain responsible for language acquisition and refuse the separation of semantic representations from the rest of cognition (Lakoff and Johnson, 1980). They claim that even

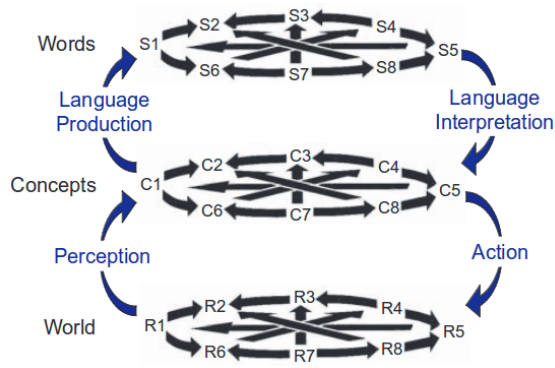


Fig. 1: General scheme of the relationship between the (multimodal) network of concepts and (symbolic) word networks, consistent with grounded cognition paradigm (from Roy 2008).

abstract concepts are grounded metaphorically in embodied and situated knowledge (for instance, love can be understood as eating, consuming the beloved person). Cognitive simulation theories focus on the role of modal simulation, situated action and bodily states in the grounding of cognitive processes (Barsalou, 1999). Social simulation theories propose that the understanding of mental states in other people requires simulations of our own mind (Goldman, 2006) and typically it requires the activation of the mirror neuron system (Rizzolatti and Craighero, 2004).

2 Psychology and linguistics

Within the relevant disciplines, perhaps most information and knowledge related to abstraction has been accumulated in cognitive psychology and psycholinguistics. The focus has been put on abstract words and conceptual knowledge regarding abstract concepts as categories. The relationship between concepts and words is sketched in Fig. 1, consistently with grounded cognition paradigm. Concepts are mental constructs created through sensory-motor interaction. On the other hand, words are surface forms (phonological or orthographic) embedded in a concrete language, arbitrarily linked to their meanings.

Three major classical *theories of categorization* have been proposed: In *rule-based* theory, members of a category share common (perceptual) properties (e.g. colour, shape, etc.) and the membership for a category is based on satisfying established rules that permit to verify the common properties of the category. Hence, categories have strict boundaries which only works well for some categories, e.g. mathematical objects, not so much for natural categories (Bruner and Austin, 1986).

In *prototype-based* theory, categories are represented by “prototype” stimuli, which are used for judging the membership of other items. This approach as-

sumes a more continuous way of categorization with less strict boundaries between categories which works well for natural categories (Rosch, 1973).

In *exemplar-based* theory, concepts are represented by the exemplars of the categories stored in the memory. A new item is classified as a member of the category if it is similar to one of the stored exemplars in that category (Nosofsky et al., 1992).

2.1 Meanings of abstraction in psychology

In cognitive psychology, Barsalou (2003) suggested that abstraction is not a unitary concept but has several meanings:

1. *Categorical knowledge*, which has been abstracted from experience (e.g. ‘chairs’). Various accounts of knowledge are comfortable with this sense, including rule-based, prototype, exemplar, and connectionist theories.
2. *Behavioural ability to generalize* (without exceptions) across category members (e.g. using generic claims “cats have fur”, or quantifications, such as “some mammals swim”).
3. *Summary representations* of category members in long-term memory (in some theories). In this sense, it is not necessary to produce the behavioural abstractions. For example, exemplar models do not store summary representations and produce behavioural abstractions by scanning and summarizing exemplars online (Hintzman, 1986).
4. *Schematic representations* of categories in memory, where “schematic” refers to summary representations being sparser than exemplars (due to extracting the critical properties of a category’s exemplars and discarding irrelevant properties).
5. *Flexible representations* as a result of flexible application of summary representations to many different tasks (e.g. categorization, inference, language comprehension, reasoning). From this perspective, increasing abstractness allows a representation to become increasingly flexible.
6. *Abstract concepts* that are typically detached from physical entities and more associated with mental events (Paivio, 1986; Barsalou, 1999; Wiemer-Hastings et al., 2001).

Barsalou (2003) claims that meanings 3, 4 and 5 are controversial, and focuses on one of them, summary representations, within his theory of perceptual symbol systems (Barsalou, 1999). In this paper, we focus on senses 1 and 6 that, despite not being controversial, still pose a challenge for (computational) cognitive science.

Categorical knowledge refers to concepts that can be organized *hierarchically*. This applies to concepts expressed by nouns but also by action words.

2.2 Abstraction in object words

Regarding nouns, three levels of categorization have been proposed: subordinate, basic and superordinate levels (e.g. ‘a rocking chair’/‘chair’/‘furniture’). *Subordinate* level categories are characterized by a low degree of generality and by clearly identifiable, detailed and specific features. *Basic* level provides (ecologically) the most relevant conceptual information. *Superordinate* level typically implies a high degree of generality and allows to store general information. As a result, subordinate categories are more concrete than basic categories, which in turn are more concrete than superordinate categories. Interestingly, despite the order of increasing abstractness, the acquisition of knowledge in children, does not proceed bottom up. Instead, it starts from the basic level, moving to subordinate and superordinate levels of categorization in parallel Bloom (2000).

2.3 Abstraction in action words

The organization of action words is also hierarchical. The most concrete action words may correspond to motor primitives, such as ‘push’, ‘pull’, ‘grasp’, ‘release’, etc. More abstract action words, related to physical actions, require the concatenation of motor primitives; e.g. we could consider ‘keep’=‘grasp’+‘stop’, or ‘give’=‘carry’+‘release’ (Stramandinoli et al., 2011). Action verbs can also be differentiated in their level of concreteness and motor modality; e.g. ‘push’ is uniquely linked with the action of pushing by using the hand, while ‘give’ implies multiple motor instances of the process of passing an object by using one hand, two hands, mouth, etc. (Cangelosi and Schlesinger, 2014).

The above examples also apply to abstract actions. For example, we can grasp or keep an idea, wrap up a meeting, etc. (as explained by cognitive linguistics). Many action verbs are inherently abstract and refer to mental operations. For instance, in the educational setting, the well-known revised Bloom’s taxonomy categorizes action verbs in cognitive, affective and psychomotor domains (Anderson and Krathwohl, 2001). For instance, in the cognitive domain, the verbs are categorized hierarchically (in ascending order) according to the abilities they involve: remembering, understanding, applying, analyzing, evaluating and creating. In general, verbs are clearly the most complex word category with verb semantics posing a challenge for cognitive science (Levin, 1993).

2.4 Differences between concrete and abstract words

Concrete and abstract words can be differentiated according to the following factors (Kousta et al., 2011):¹

¹Altarriba et al. (1999) proposed that words referring to emotions should be categorized as distinct from concrete and abstract words,

- *Perceivability* As opposed to concrete, abstract words do not have a physical referent and hence refer to entities that are distant from immediate perception, and typically represent entities that are not spatially constrained (e.g. ‘truth’, ‘happiness’).
- *Imageability and context availability* According to the dual-coding theory (Paivio, 1986), concrete concepts, activating both verbal and non-verbal systems, require a major involvement of memory. On the contrary, abstract concepts are represented in the verbal system only, with inferior involvement of memory (Barsalou, 2008). In addition, they evoke less imagery (Wiemer-Hastings and Xu, 2005), and activate a narrower contextual verbal support than concrete concepts, as suggested in the context-availability theory (Schwanenflugel, 1991). These can be reasons why acquisition of abstract concepts is more complex than that of concrete concepts.
- *Hierarchical categorization* Conceptual knowledge can be organized in categories hierarchically, as already mentioned in Section 2.2. This works well for concrete objects, but even there the concepts at the superordinate level can become very abstract, unifying subcategories that may not share any perceptual features (e.g. ‘game’).
- *Mode of Acquisition* It has been shown in experiments with elementary school children that MOA ratings change with the school age, shifting gradually from mainly perceptual to mainly linguistic MOA (Wauters et al., 2003).

Abstract concepts pose a bigger challenge for embodied (grounded) theories of cognition, because the grounding of abstract concepts, deprived of their physical referents, seems less straightforward than that of concrete concepts. Symbolic cognition treats all concepts as symbols, so the burden is shifted to accounting how all concepts acquire their meanings (Harnad, 1990).

Abstract concepts are also hierarchically organized (e.g. ‘love’ is less abstract than ‘democracy’), depending on how well they can be grounded. Even abstract concepts appear to depend heavily on situations and situated action (Schwanenflugel, 1991). The processing of abstract concepts is facilitated when a background situation contextualizes it (Barsalou and Wiemer-Hastings, 2005). We can also add that for concrete objects, one usually abstracts over individuals, whereas for abstract objects one abstracts over situations or events.

Categorizing entities is a useful approach in science, but consistently with cognitive theories, it is more accurate to think of the distinction between concrete and abstract words as that of a *concreteness–abstractness continuum*, along which all entities can be varied. It

because these three categories received different ratings in term of concreteness, imageability and context availability.

also applies to cognitive theories that range from purely symbolic (disembodied) accounts to purely embodied ones (Wilson, 2002). In addition, some have argued that both disembodied and embodied approaches are needed for complete account how humans acquire knowledge (Louwerse, 2010; Dove, 2011).

On the sentence level, comprehension and production processes focus, within the grounded cognition paradigm, mainly on the roles of *mental simulation* (e.g. Barsalou 1999; Decety and Grèzes 2006; Goldman 2006). Simulation is the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind. The question is how even very abstract sentences can be simulated. Zwaan (2015) argues that sensorimotor and symbolic representations mutually constrain each other in discourse comprehension, and proposes that abstract concepts serve two roles in this process. They serve as pointers in memory, used (1) *cataphorically* to integrate upcoming information into a sensorimotor simulation, or (2) *anaphorically* to integrate previously presented information into a sensorimotor simulation. In either case, he concludes, the sensorimotor representation is a specific instantiation of the abstract concept.

3 Neural perspective

Neurophysiological and neuropsychological evidence suggests hierarchical organization of knowledge (long-term memory) in the brain that can be observed mainly over the cortex (Fuster, 2009). Fuster's theory of memory organization is based on the existence of bidirectionally connected executive and perceptual memories, interacting at multiple levels, both forming the so-called *cognits*, formed by distributed neural networks. Cognits are items of knowledge, hierarchically organized in terms of semantic abstraction and complexity. As shown in Figure 2, frontal and posterior cortices provide different spatial layout of organization. Whereas in the frontal lobe, the apex of the hierarchy rests at the very anterior part of dorsolateral prefrontal cortex, the top of hierarchy in the posterior cortex rests "inside" the lobes. Fuster's theory departs from the idea of a memory as a passive storage, and instead sees the memory as operating in perception–action cycles (Fuster, 2004), which is a core concept in biology (Uexküll, 1926).²

3.1 Posterior cortex

In the temporal cortex, the increasing abstraction can be related to the growing invariance of the neurons in response to various perceptual stimuli. The visual system of mammals is organized hierarchically in terms of feature extraction (from the simplest to the more complex

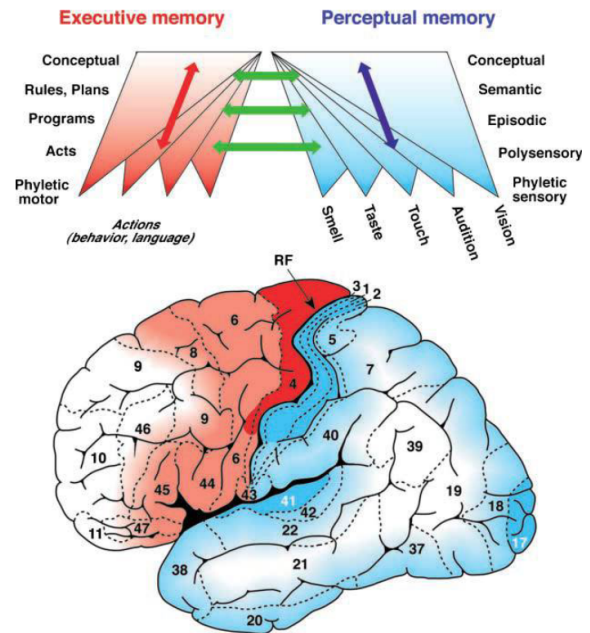


Fig. 2: General scheme of the hierarchical organization of memory in the lateral cerebral cortex of the left hemisphere (Fuster, 2009). Lighter shade of colors indicates the direction in the hierarchy towards more abstract memory and knowledge.

features), which is associated with increased radius of the receptive field of neurons (via afferent connections) towards the higher (more abstract) layers. On top of the pyramid stands the area IT (inferior temporal), responsible for invariant object recognition with respect to position, rotation, and scale, and in the case of biological objects also deformation (Jellema and Perrett, 2006). Likewise, area STS (Superior Temporal Sulcus) has its division in the context of the recognition of biological motion. An interesting property of STS is that it contains many neurons that are sensitive to viewpoint from which the object is observed (e.g. front view, side view, etc.) in its posterior part (STSp), but also neurons that are invariant to it, in anterior part (STSa).

Parietal cortex also demonstrates abstraction properties. For instance, intraparietal sulcus has been argued to be the seat of (abstract) numerical cognition (Dehaene et al., 1998) (but see the challenging view in Kadosh and Walsh 2009). Among other things, posterior parietal cortex is known to be involved in spatial cognition (the superior parietal lobule) and in non-spatial cognition (the inferior parietal lobule). The latter is typically related to abstract information processing. For instance, Yamazaki et al. (2009) proposed several mechanisms on non-spatial representations processed by the posterior parietal cortex, ranging from perceptual space (low abstraction), via temporal and social spaces to conceptual space (high abstraction), with corresponding brain areas involved (collected from previous papers).

²This is consistent with a psychological model of memory which is postulated to subserve intelligent action Glenberg (1997).

3.2 Frontal cortex

Parietal cortex has bidirectional connections to the frontal cortex, which is crucial for sensory-motor integration. Various levels of abstraction are subserved by corresponding mirror neurons operating simultaneously at various levels of a perceptual hierarchy (e.g. perspective-dependent neurons as opposed to goal-coding mirror neurons, that are insensitive to motor trajectories). Mirror neuron system has been labeled as the action-observation-action-execution matching circuit, supporting the direct matching hypothesis (Rizzolatti and Craighero, 2004). This circuit is enriched by its connections (from the parietal cortex) to the temporal cortex. Namely, a part of the focus area F5 (F5c) is connected with STSp through PFG (parietal frontal gyrus) forming a perspective variant path. Another part of F5 (F5a) is also connected with STSa through AIP, forming an invariant path emphasizing the actor and the object acted upon, rather than the viewpoint from which it is observed (Nelissen et al., 2011).

The frontal lobes are most expanded in humans, being hierarchically organized, all the way from the primary motor cortex towards the prefrontal cortex (PFC). Highest in the pyramid stands the apex of the dorso-lateral PFC. It is noteworthy that different parts of the frontal cortex have a corresponding structure (with connections to the thalamus and some of the nuclei in the basal ganglia, which are involved in action selection (O'Reilly et al., 2012)). It thus appears that similar (or identical) neural mechanisms operate at different levels of the hierarchy (abstraction).

Abstraction in PFC seems to comply with several different types:

- *Temporal abstraction* implies integration of behavior (Koechlin and Hyafil, 2007; O'Reilly, 2010), maintaining the organization of (sub)actions with increasing temporal spans. For example, the (more abstract) goal to make coffee can be sequentially decomposed to several subgoals (get a cup, pour water in a machine, etc.), which in turn can eventually be decomposed to a set of primitive actions.
- *Policy abstraction* (Badre and D'Esposito, 2007) implies the representation of a task, at any given moment of performance, as a set of choices over lower-level tasks. For instance, a certain action can be executed in various (motor) ways, achieving the same goal.
- *State abstraction* Christoff and Keramiatian (2007) occurs when the subject treats non-identical stimuli or situations as equivalent. For instance, we ignore the font type during reading.

In sum, the brain seems to have two major abstracting pathways with revealed neural organization: one located posterior for (physical) object recognition,

and the other one in frontal lobes, corresponding to abstracting in execution and planning of actions. Probably the third pathway, corresponding to hierarchical organization of abstract objects (words) could also be searched for in the brain. The review and meta analysis of the fMRI literature provides a frame for this search (Binder et al., 2009). It was concluded that the semantic memory (covering the processing of both concrete and abstract words) is distributed in a left-lateralized network consisting of 7 distinct regions.³

Similar meta analysis of fMRI and PET studies (Wang et al., 2010) indicated consistent and meaningful differences in neural representations. Roughly speaking, abstract concepts elicit greater activity in the frontal areas, while concrete concepts elicit greater activity in the posterior brain areas. These results suggest greater engagement of the verbal system for processing of abstract concepts and greater engagement of the perceptual system for processing of concrete concepts.

4 Computational aspects

The role of computational modeling is increasingly growing⁴, promising a huge potential for advancing knowledge in the interdisciplinary cognitive science (McClelland, 2009). In addition, the importance of computational modeling can be considered an indispensable ingredient allowing the mechanistic, neurally constrained understanding of the mind, extending the horizons of cognitive science (Farkaš, 2012).

4.1 Towards deciphering the brain code

The brain representations in IT area (mentioned in Section 3.1) seem quite reliable despite ever-present noise because the technology already allows their reading from the brain (Hung et al., 2005), using single-cell recordings in monkeys. However, even the non-invasive imaging methods (e.g. fMRI) provide means for deciphering the brain.

Mitchell et al. (2008) reported about a computational model that could not only be used for interpreting various spatial patterns of neural activation corresponding to different semantic categories (expressed by concrete nouns), but also for predicting (previously unrecorded) fMRI patterns for novel concrete nouns (via generalization). This was possible due to linking the neural data to the statistics of very large text corpora.

A nice global picture of the growing abstraction in the brain has been recently proposed as a result of a large-scale analysis of fMRI data (~17,000 experiments and ~1/4 of fMRI literature), i.e. the technology, that

³These are posterior inferior parietal lobe, middle temporal gyrus, fusiform and parahippocampal gyri, dorsomedial PFC, inferior frontal gyrus, ventromedial PFC, and posterior cingulate gyrus.

⁴despite the dominant role of psychology (Gentner, 2010)

has increasingly become an important source of information over the last decades (Taylor et al., 2015). In the paper, the hypothesis was tested, using formal methods based on a new cortical graph metrics (network depth), that regions deeper in the brain (i.e. more remote from the sensory inputs) represent more abstract functions. Data-driven analyses defined a hierarchically ordered connectome, revealing a related continuum of cognitive function. The authors concluded that progressive functional abstraction over network depth may be a fundamental feature of brains.

It can be expected that the number of papers of this sort will grow, revealing the patterns represented by the brain. Most probably, deciphering abstract concepts from the brain will remain a challenge.

4.2 Neural network models

So far, four major frameworks have been proposed for representing conceptual knowledge (symbolic, connectionist, dynamic and probabilistic) (McClelland, 2009). Each of the frameworks has to be able to incorporate abstraction, in order to be able to account for the spectrum of cognitive processes, ranging from sensory-motor behaviors to abstract thoughts. Believing that connectionism has the best potential in this regard (Farkaš, 2011, 2013), we will henceforth focus on computational aspects of abstraction from the connectionist perspective.

Neural networks perspective involves the idea of brain operation as performing a cascade of nonlinear transformations of patterns across layers. These transformations go in both directions (i.e. also backwards), reflecting the fact that most of the connections between brain areas are bidirectional. Since their breakthrough a decade ago (Hinton et al., 2006), *deep neural networks* (DNNs) have become the influential approach in connectionist modeling, fueled by very successful applications in various tasks, such as object recognition/classification, speech recognition and natural language processing (Schmidhuber, 2015). DNNs lend themselves nicely to hierarchical modeling (see Fig. 3), since at each layer the features with increasing abstraction are formed during training.⁵ Visual system of mammals is also known to have deep hierarchical structure (Felleman and Essen, 1991), starting at the retina, through the cascade of feature detecting layers with an increasing complexity, all the way up to IT cortex where (physical) object recognition takes place.

It is important to realize that this pathway only applies to concrete (imageable) objects which can be still be considered a rather low level of abstraction, compared to abstract objects (entities). In case of abstract objects, probably a different neural pathway has to be sought, because these do not have direct reference in

⁵Typically in classification tasks, most layers are trained in unsupervised way, and are combined with supervised training at the top layer (Hinton et al., 2006).

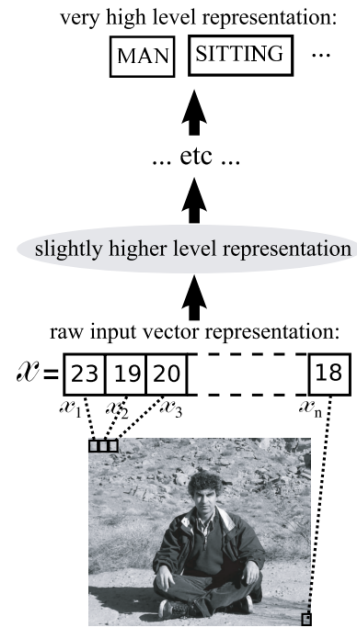


Fig. 3: Picture illustrating the “distance” between the raw data and high-level description of the image (Bengio, 2009). The established mapping will involve a number of intermediate levels of representation with an increasing abstraction.

the environment. Even though abstract concept representations can be triggered by visual inputs (visual context), they will be more linked to language. Linguistic context can hence as well serve as the cue. Representations of abstract objects are usually modeled not as deep networks but as widely distributed associative networks spread over distant cortical (and subcortical areas) (Pulvermüller, 1999). Shallice and Cooper (2013) suggest that networks for abstract words differ from those for concrete words, and depend more on frontal regions (left PFC or left inferior frontal gyrus) that are known to be involved in syntactic processing (and hypothesized as crucial for understanding abstract words).

5 Conclusion

Our understanding of abstraction is far from complete, regardless of the perspective taken. In cognitive psychology and linguistics, good evidence has been accumulated about how concrete and abstracts objects depend on the environment and how they differ regarding their properties and the role of language in their acquisition and use (more important for abstract concepts). Better understanding of semantic organization of content words (referring to concepts) in the brain could provide important material to be integrated in cognitively and biologically inspired modeling of this phenomenon. Organization of semantic knowledge on the sentence level and in discourse is yet more elusive.

Acknowledgment

This work was supported by VEGA project 1/0898/14 and KEGA project 017UK-4/2016.

References

- Altarriba, J., Bauer, L. a Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods*, 31(4):578–602.
- Anderson, L.W., a Krathwohl (Eds.). (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Badre, D. a D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, 19:2082–2099.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(04):577–660.
- Barsalou, L. (2003). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1435):1177–87.
- Barsalou, L. (2008). Grounded cognition. *Annual Reviews of Psychology*, 59:617–645.
- Barsalou, L. a Wiemer-Hastings, K. (2005). *Situating abstract concepts*, pp. 129–163. Cambridge University Press.
- Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and Trends in Machine Learning*, 2(1):1–127.
- Binder, J., Desai, R., Graves, W. a Conant, L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19:2767–2796.
- Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press, Cambridge, MA.
- Borghi, A., Flumini, A., Cimatti, F., Marocco, D. a Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Frontiers in Psychology*.
- Bruner, J. a Austin, G. (1986). *A Study of Thinking*. Transaction Books.
- Cangelosi, A. a Schlesinger, M. (2014). *Developmental Robotics: From Babies to Robots*. MIT Press, Cambridge MA.
- Christoff, K. a Keramatian, K. (2007). Abstraction of mental representations: theoretical considerations and neuroscientific evidence. In *The Neuroscience of Rule-Guided Behavior*, pp. 107–127. Oxford University Press.
- Decety, J. a Grèzes, J. (2006). The power of simulation: imagining one's own and other's behavior. *Brain Research*, 1079:4–14.
- Dehaene, S., Dehaene-Lambertz, G. a Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, 21:355–361.
- Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Frontiers in Psychology*, 1(242).
- Farkaš, I. (2012). Indispensability of computational modeling in cognitive science. *Journal of Cognitive Science*, 13(12):401–435.
- Farkaš, I. (2011). Connectionism embraced by computational cognitive science (in Slovak). In *Umelá inteligencia a kognitívna veda III*, pp. 19–62. Vydavateľstvo STU, Bratislava.
- Farkaš, I. (2013). Is representational pluralism in cognitive science inevitable? (in Slovak). In *Kognitívni veda a umelý život*, pp. 107–114. Slezská univerzita v Opavě.
- Felleman, D. a Essen, D. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47.
- Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.
- Fuster, J. (2004). Upper processing stages of the perception–action cycle. *Trends in Cognitive Science*, 8:143–145.
- Fuster, J. (2009). Cortex and memory: Emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21(11):2047–2072.
- Gentner, D. (2010). Psychology in cognitive science: 1978–2038. *Topics in Cognitive Science*, 2:328–344.
- Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences*, 20:1–55.
- Goldman, A. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.
- Harnad, S. (1990). The symbol grounding problem. *Physica D*, pp. 335–346.
- Hinton, G., Osindero, S. a Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554.
- Hintzman, D. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, 93:411–428.
- Hung, C., Kreiman, G., Poggio, T. a DiCarlo, J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310:863–866.
- Jellema, T. a Perrett, D. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia*, 44:1535–1546.
- Kadosh, R. a Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, 32:313–373.

- Koechlin, E. a Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, 318:594–598.
- Kousta, S., Vigliocco, G., Vinson, D., Andrews, M. a Campo, E. D. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, 140(1):14–34.
- Lakoff, G. a Johnson, M. (1980). *Metaphors We Live By*. University Chicago Press, Chicago.
- Levin, B. (1993). *English Verb Classes and Alternations*. Chicago University Press, Chicago.
- Louwerse, M. (2010). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, pp. 1–30.
- McClelland, J. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, 1(1):11–38.
- Mitchell, T. et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195.
- Mussa-Ivaldi, F. a Bizzi, E. (2000). Motor learning through the combination of primitives. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404):1755–1769.
- Nelissen, K. et al. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, 31(10):3743–3756.
- Nosofsky, R., Kruschke, J. a McKinley, S. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2):211–233.
- O’Reilly, R. (2010). The what and how of prefrontal cortical organization. *Trends in Neurosciences*, 33(4):355–361.
- O’Reilly, R., Munakata, Y., Frank, M., Hazy, T. a contributors (2012). *Computational Cognitive Neuroscience*. Wiki book, 2. vyd.
- Paivio, A. (1986). *Mental Representation: A Dual Coding Approach*. Oxford University Press, Oxford.
- Pulvermüller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, 22:253–336.
- Rizzolatti, G. and Craighero, L. (2004). The mirror neuron system. *Annual Reviews of Neuroscience*, 27:169–192.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3):328–350.
- Roy, D. (2008). A mechanistic model of three facets of meaning. In *Symbols and Embodiment: Debates on Meaning and Cognition*, pp. 195–222. Oxford University Press.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117.
- Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? In *The Psychology of Word Meanings*, pp. 223–50. Lawrence Erlbaum.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–424.
- Shallice, T. a Cooper, R. (2013). Is there a semantic system for abstract words? *Frontiers in Human Neuroscience*, 7:1–10.
- Stramandinoli, F., Cangelosi, A. a Marocco, D. (2011). Towards the grounding of abstract words: A neural network model for cognitive robots. In *Proc. of the Int. Joint Conf. on Neural Networks*, pp. 467–474.
- Taylor, P., Hobbs, J., Burrioni, J. a Siegelmann, H. (2015). The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. *Scientific Reports*, 5, doi:10.1038/srep18112
- Thagard, P. (2005). *Mind: An Introduction to Cognitive Science*. MIT Press, second. vyd.
- Uexküll, J. (1926). *Theoretical biology*. Harcourt, Brace and Co., New York.
- Wang, J., Conder, J., Blitzer, D. a Shinkareva, S. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, 31(10):1459–1468.
- Wauters, L., Tellings, A., van Bon, W. a van Haften, A. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, 24(3):385–406.
- Wiemer-Hastings, K., Krug, J. a Xu, X. (2001). Imagery, context availability, contextual constraint, and abstractness. In *Proc. of the 23rd Annual Conf. of the Cognitive Science Society*, pp. 1134–1139.
- Wiemer-Hastings, K. a Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29(5):719–736.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625/636.
- Wintermute, S. (2012). Imagery in cognitive architecture: Representation and control at multiple levels of abstraction. *Cognitive Systems Research*, 19–20:1–29.
- Yamazaki, Y., Hashimoto, T. a Iriki, A. (2009). The posterior parietal cortex and non-spatial cognition. *F1000 Biology Reports*, 1:1–6.
- Zwaan, R. (2015). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin and Review*, doi:10.3758/s13423-015-0864-x

Modelling subjective experience¹

Silvia Gálíková

Katedra filozofie, Filozofická fakulta
Trnavskej Univerzity v Trnave,
Hornopotočná 23, 91843 Trnava
silvia@libris.sk

Abstract

In contemporary cognitive science, philosophy of mind, computer science and neuroscience, there has been ongoing intense research in the study of cognitive states such as problem solving, planning, decision making, memory, mental imagery etc. Due to the heterogeneity of types and kinds of subjective states, absence of relevant methods of inquiry, modelling of states of subjective experience became an integral part of experimental research rather late. One of the most prevalent obstacles rested on inability to grasp the nature of inner experience from an objective perspective. In my presentation I intend to outline selected representative models of subjective experience from phenomenological, neuroscientific, cognitive and conceptual standpoints. My aim is to point out, especially, the explanatory force and implications arising from the „*phenomenal self-model*“ of T. Metzinger and the potential of „meeting of minds“ model based on „conceptual spaces“ of P. Gärdenfors.

1 Introduction

What is the relation between phenomenal consciousness and the self? In what sense can the “self” be a real object in the world? What are we talking about when we speak

of conscious experience, the self, an inner mental world? What is the basic structure and conceptual framework of self-experience? Does the self have an experiential reality or is it nothing but a theoretical fiction? Is there any use of the self in our conscious lives? What would a theory of subjective experience look like? These questions present just a small fraction from the multitude of open questions related to the phenomenon of the self and inner experience. Difficulties in answering these persisting questions are closely related to a “terminological fog” (which is) spread over the study of the mind and consciousness. The following text reflects recent effort to model types, structure, origin, mechanisms and functions of states of inner experience. Distinctive problem has been tied with the birth and functioning of conscious states and their „bearer“ – the conscious self. In contemporary research on the nature of our experience we have at our disposal a variety of models, f.e. structural, symbolic, connectionist, artificial neural networks, semantic, neurocomputational, cognitive, linguistic, multiagents etc. Questions and problems derive from inconsistencies existing among individual models which brings difficulties in their comparison and evaluation. Heterogeneity of criteria and validity conditions are present in often radically different or contradicting starting points and following expectations of a study of subjective experience. The core of the ongoing controversies lies in the absence of a common, at least working, characteristic of phenomena of experience, account on conscious/unconscious

¹ In the present text I will use the concepts of self, subjective states, inner experience in a broad sense – as umbrella terms for both conscious and unconscious states of experience.

distinction and the status of self. The experimental study of normal subjects in laboratory and natural settings, together with clinical work on the impairments and disturbances of the self, have improved significantly our knowledge on the complexity of the states of the self. In the present text I argue that in order to give credence to subjective experience it is necessary to adopt a naturalistic perspective. This requires to treat phenomena of inner experience as natural as any other phenomena – a subject to scientific methodology and aims of inquiry. It is crucial to highlight the fact that in studying and explaining subjective experience there is no difference in principle with respect to such natural phenomena as life, behaviour or global warming.

1.1 What? of the “inner”

The seemingly unproblematic existence of inner experience in everyday experience, our thoughts, beliefs, feelings has become a deep theoretical problem. Past decades brought a significant revival and interest into the nature of our inner experience. New tools such as neurotechnologies, brain scanning techniques, illuminating ideas and theories, experimental and clinical findings greatly enhanced our knowledge of the workings of our minds. It seems that we have never been so close to reveal the apparent mysteries of our inner lives. Paradoxically, this is far from being a unified picture of current state of art in mind and consciousness studies. Interestingly enough, in spite of controversies regarding the study and explanation of inner subjective life, as indicated above, scholars seem to agree upon the following:

-The mind and self are embodied, a picture of disembodied reason, the self, consciousness existing independently of

bodily structures is either a categorical mistake, a myth or both

- The workings of mind are largely unconscious
- Cognition is both embodied and embedded – people as cognitive agents live and act in the world
- The reality of the self has its origin in the way we experience ourselves from the inside
- Experimental and clinical findings are highly relevant to the study of the phenomena of the self and subjective experience

Thus, at first sight, it appears that representatives of a variety of conceptions treat self and states of experience as natural phenomena, open to scientific research and explanation. The most prevalent issue of disagreement is related to the question of the starting point, followed by relevancy of used methods and aims of inquiry. It is unclear what kind of phenomenon philosophers, scientists and theorists aim to study, explain and formulate a theory of? Thus, an urge for interdisciplinarity often remains just a proclamation. Let me illustrate the point on the question of the starting point. The existence and function of the subjective experience has been approached from a number of different perspectives towards *reality*. Even if naturalistic approaches can be considered as prevailing there is a crucial difference between theoretical conceptions which approach the nature of inner experience as principally *irreducible* and those accounts which consider our inner life reducible as any other natural phenomena. The traditional *naturalistic* and *phenomenological* (transcendental) “distinction” reappears in new forms. Naturalistic attitude presupposes the existence of the natural world, objective physical space-time. Within this framework creatures with their conscious experiences are considered as a part of the

physical world approachable by a third-person perspective. Followers of the naturalistic-reductionist attitude offer interesting models and theories, for example P. Churchland, D. Dennett, A. Damasio, R. Llinás, D. Edelman etc. Recently, a number of phenomenologists realized that in order to go hand in hand with contemporary science in the study of the mind it is necessary to embrace empirical findings. F. Varela, as a representative of the phenomenological attitude claims that: “The phenomenological approach starts from the *irreducible* nature of conscious experience. Lived experience is where we start from and where all must link back to, like a guiding thread” (Varela, 1996, 6). For Varela, phenomenology’s foundational project is the re-discovery of the primacy of human experience and its direct, lived quality. He has proposed a philosophical account – *neurophenomenology* – bridging the phenomenal domain and empirical observations of neuronal activity. The phenomenal term stands as an explicit term *directly* linked to experience by a rigorous examination based on reduction, invariance and intersubjective communication (Varela, 1996, 21). Varela admits a fundamental difference between notion of reduction in phenomenology and science. The notion of *phenomenological reduction* (as a part of the overall method) based on “bracketing” the world of everyday experience, appears right at the beginning of the inquiry. At the same time, *irreducibility* of conscious experience is used in a strict ontological meaning. *Theoretical reduction* in empirical science has been practised during the process of testifying the research hypothesis. In traditional terms it aims at reducing complex interactions and entities to the sum of their constituent parts, in order to make them easier to study. A possible outcome of scientific reductionist strategy – *ontological reduction* - comes close to the end of the investigation

or may not come at all. For D. Zahavi (2005) crucial phenomenological findings have to go hand in hand with contemporary theoretical and ongoing experimental research in a variety of disciplines. The self possesses an experiential reality that means that it is linked to the first-person perspective or first-person *givenness* of the experiential phenomena (Zahavi, 2005, 106). The self is not an ineffable transcendental precondition or a mere social or narrative construct, it is an integral part of our conscious life with an immediate experiential reality. The aim of phenomenological analysis is to describe and explain experiential structures, or more precisely, the fundamental structural principle that, like space, enforces the nature of what is given to us as the contents of that experience. So, phenomenology seeks not so much to understand how physicality opens up the experience of the self, the world and others, but rather aims to specify the *mechanisms* that explain how cognition is grounded in, and deeply constrained by, the bodily nature of cognitive agency. The general characterisation of embodied cognition provides the basis for the embodiment thesis², meaning that many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing.

² G. Lakoff and M. Johnson (1999) identified embodiment strongly with bodies (sensorimotor experience) and neuronal activity of the brain. That the mind is embodied and its workings are largely unconscious has been analysed with the help of the concept of *cognitive unconscious* - a realm of thought that is completely inaccessible to direct conscious introspection, such as automatic cognitive operations or implicit knowledge. Conceptual systems and our capacity for thought are shaped by the nature of our brains, bodies, and bodily interactions.

Recent discussions have revealed even more diverse meanings of the term self. G. Strawson, who considers himself as a naturalist, argues that: when I talk about selves I mean selves as classically conceived: “internal mental things, internal mental presences” (Strawson, 2010). Whatever a self is, it is certainly a subject of experience and it is not a human being considered as a whole. So, according to Strawson, the problem about the existence of a self arrives from our everyday self-experience. It is a self described as: a) a subject of experience, a conscious feeler and thinker, b) a thing, c) a mental thing, d) a persisting thing, e) an agent, or f) something that has a personality.

1.2 Lessons from simulation studies

Highly debated experimental findings on the status of the self, self-agency and self-consciousness relate to T. Metzinger’s now-classic experiment known as the “*the rubber-hand illusion*”. Participants of this experiment observed a rubber hand lying on the desk in front of them. At the same time, their own corresponding hand has been concealed from their view by a screen. The visible rubber hand and the unseen hand were synchronously stroked with a probe. After a certain period of time (cca 60 seconds), the rubber-hand illusion emerges. Participants experienced the rubber hand as their own and felt repeated strokes in the rubber hand. They also felt a “virtual arm” – a connection from their shoulder to the fake hand on the table in front of them. There have been more experiments that have simulated an experience of not only a part of the body, but the body as a whole. This is the case of an *out-of-body experience* a condition during which people experience a sense of being outside their bodies, looking at themselves for above. Scientists have been recently able to simulate this phenomenon using a rather simple technique. The experiment of creating an *out-of-body experience* was conducted by the neuroscientist, O. Blanke (2012). During

the experiment, the test participant would wear virtual reality goggles while standing in an empty room. A camera located behind the participant projected an image – which appeared as if it were six feet in front of the participant – within the goggles. O. Blanke would then poke the participant’s back for one minute with a stick, this action also being visible within the goggles. By varying the delay of the images shown in the goggles, Blanke was able to detect certain differences in the experience. When the poking was synchronous, tested people reportedly had a sense of being momentarily within the illusory body. When the pokes were not synchronous, this illusion did not occur. The experimental induction of out-of-body experiences outlines the unique method by which the illusion was created and also brings interesting implications for further study. *Out-of-body experiences* have been reported in clinical conditions where brain function is compromised, such as a stroke, epilepsy and drug abuse. They have also been reported in association with traumatic experiences such as car accidents.

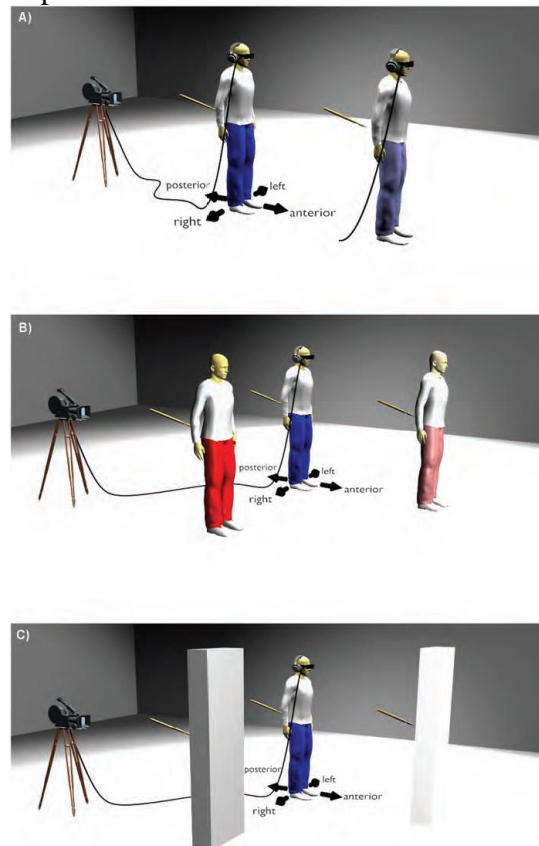


Figure 1: (A) Participant (in dark blue trousers) sees through a HMD his own virtual body (light blue trousers) in 3D, standing 2m in front of him and being stroked synchronously or asynchronously at the participant's back. In other conditions (Study II) the participant sees either (B) a virtual fake body (light red trousers) or (C) a virtual non-corporeal object (light gray) being stroked synchronously or asynchronously at the back. Dark colors indicate the actual location of the physical body/object, whereas light colors represent the virtual body/object seen on the HMD. Illustration by Martin Boyer, published in Science.

2 Egological or non-egological self?

A crucial point of departure between naturalists reductionalist and naturalists nonreductionalists on the nature of the self has been reflected within contemporary debate on *egological* and *non-egological* accounts of the self. For the followers of egological theories, it is in general a conceptual and experiential truth that any episode of experiencing necessarily includes a subject of experience. Non-egological theories, on the contrary, deny that every experience is for a subject. A prominent representative of the first approach is D. Zahavi and of the second neuroscientist T. Metzinger who has proposed a theory based on a kind of *no-ownership* view. For Metzinger the phenomenological notion of the self has no explanatory value and in this sense nothing exists “behind” the concept itself. He considers *Ego tunnel* as the central metaphor for conscious experience. Conscious experience is like a tunnel: what we see and hear or we feel, smell and taste is only a small fraction of what actually exists out there. Our conscious model of reality is a low-dimensional projection of the much richer physical reality surrounding us. Our sensory organs are limited: They evolved for reasons of survival, not for depicting the enormous wealth and richness of reality in all its depth. Therefore the ongoing process of conscious experience is not so much an image of reality as a tunnel *through* reality (Metzinger, 2010, 6). The content of the phenomenal, what is known experientially,

through the ways things appear to us - self-model – the conscious model of the organism as a whole that is activated by the brain.

2.1 From self model³ to phenomenal self-model

The *self-model* is the central concept in the theory of consciousness called the *self-model theory of subjectivity* (SMT). This concept comprises experiences of ownership, of first person perspective, and of a long-term unity of beliefs and attitudes (instantiated in the prefrontal cortex). Metzinger's theory has two core contents, the phenomenal self-model (PSM) and the phenomenal model of the intentionality relation (PMIR). The PSM is an entity that “actually exists, not only as a distinct theoretical entity but something that can be empirically discovered in the future - for instance, as a specific stage of the global neural dynamics in the human brain”. In order to explain the concept of the self - PSM has to involve three phenomenal properties: a) mineness, “a higher order property of particular forms of phenomenal content,” or the idea of ownership, b) perspectivalness, the ecological self, the immovable center of perception and c) selfhood, which is “the phenomenal target property” or the idea of the self over time. It is the property of phenomenal selfhood that plays the most important role in creating the fictional self and the first person perspective. Metzinger defines the first person perspective as the “existence of single coherent and temporally stable model of reality which is representationally centered around or on a single coherent and temporally stable phenomenal subject”. The first-person perspective can be non-conceptual and is autonomously

³ The most prominent representative of the theory of mental models Philip Johnson-Laird (1983, 1988, 1995), has introduced and developed the mental model theory predominantly to understand reasoning or perception in a specific representational framework.

active due to the constant reception of perceptual information by the brain, specifically the brainstem and hypothalamus, processes this information into representational content, namely linguistic reflections. The PSM uses this representational content to attribute phenomenal states to our perceived objects and ourselves. The data structures and transport mechanisms of the data are “transparent” so that we can introspect on our representations of perceptions, but cannot introspect on the data or mechanisms themselves. These systemic representational experiences are then connected by subjective experience to generate the phenomenal property of selfhood. Subjective experience is the result of the Phenomenal Model of Intentionality Relationship (PMIR, figure 3, 2). The PMIR is a “conscious mental model, and its content is an ongoing, episodic subject-object relation”. The model is a result of the combination of our unique set of sensory receptors that acquire input, our unique set of experiences that shape connections within the brain, and our unique positions in space that give our perception perspectivalness. The prefrontal cortex is implicated in all the functions of the human self model. The following functions all require communication with the prefrontal cortex; agency and association areas of the cortex; spatial perspectivity and the parietal lobes, unity and the temporal lobes.



Figure 2: Good Old Fashioned Intentionality (GOFI): A subject-component (S; the "mental act") is directed at an object-component (O; the "intentional object"). As O does not necessarily

exist, GOFI is a nonphysical relation.

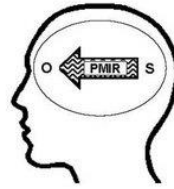


Figure 3: The Phenomenal Model of the Intentionality-Relation (PMIR): A subject-component (S; the PSM, an internal, conscious model of the system as a whole,) is phenomenally represented as directed at an object-component (O; the "intentional object"). For instance, in conscious volition O always is a goal-component, for example, an allocentric representation of a successfully terminated bodily action.

The content of the PSM is the content of the conscious self: our current bodily sensations, our present emotional situation, plus all the contents of phenomenally experienced cognitive processing. They are constituents of our PSM. Intuitively, and in a certain metaphorical sense, one could even say that you are the content of your PSM. Our self-directed thoughts operate on the current contents of our PSM, when we form thoughts about our “unconscious self” (i.e., the contents of mental self-model), these thoughts are always about a conscious representation of this “unconscious self,” one that has just been integrated into our currently active PSM. If we want to initiate a goal-directed action aimed at some aspect of ourselves, for example, brushing our hair, we need a conscious self-model to deliberately initiate these actions. Besides self-directed actions there are self-directed behaviors, for which there no need for a conscious process of goal selection – there is no PSM. Examples of unconscious, but functionally active regions in the self-model are also the body schema and low-level resonance mechanisms implemented by the mirror system in many nonlinguistic creatures (f.e. large schools of fish rapidly changing the direction when a predator is perceived just by one of them). The unconscious body schema allows us to automatically keep our bodily posture, in a

seemingly effortless manner. There are also many situations in which a previously unconscious partition of the self-model suddenly becomes conscious, that is, globally available⁴.

2.2 The Ego tunnel

Based on the self-model, phenomenal self-model and neuroscientific findings of experimental research, T. Metzinger introduced a thought provoking hypotheses on the “myth of the self”. In his book, *The Ego Tunnel*, he argues “There is no such thing as a self. Contrary to what most people believe, nobody has ever *been* or *had* a self” (Metzinger, 2010, 1).

T. Metzinger criticizes a specific philosophical conception of the self – primarily a phenomenological concept of the self as “*a given something*” postulated by a priori reasoning. Rejecting this peculiar account on the self leads some philosophers to an even more peculiar idea, that for Metzinger the self and consciousness do not exist at all. What has been highly ignored or underestimated is the *positive* aspect of Metzinger’s conception – as a testable hypothesis about how this subject becomes the center of its own world, how it transforms the appearance of a reality into a truly *subjective* phenomenon by tying it to an individual first-person perspective. From the outlined debate between naturalists reductionists and antireductionists follow rather paradoxical statements. It is not clear what scholars are speaking about, not only when they speak about the existence, but also about the non-existence of the self. The concept of non-existence has been associated with a variety of meanings, due to: a) the vagueness of the concept of the self, b) the

illusory nature of the phenomenon of the self, c) the impossibility of treating the self as a scientific explanandum, or d) an object in the surrounding world etc.

D. Dennett, for example, considers the self as an entity that is spread across the brain in time and space, and he includes both conscious and unconscious processes within the brain that are responsible for causing behavior. He is well known as a critic of the “Cartesian theatre” an idea about the place “where it all (subjective experience) happens”. Sometimes, as Dennett would put it, philosophers switch from “Cartesian dualism” to the less obscure “Cartesian materialism” where the picture of a “ghost in the machine” has been replaced by the picture of a “brain in the machine”. As Dennett writes, “those who claim to be materialists while still hanging onto the Cartesian Theatre with all its alluring imagery,” he says are trapped in “Cartesian materialism”, “the view that there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of ‘presentation’ in experience” (Dennett, 1991, 107). Dennett’s view on the self, freedom and will (Dennett, 2003) enables us to overcome the traditional free will problem, according to which free equals conscious states equals mental states, as the ultimate cause of a person’s behaviour. When S. Blackmore (2005) claims that consciousness is truly a curious illusion, she means it. At the same time, she does not say that consciousness does not exist, as her critics claim: “I mean that consciousness is not what it seems to be”. An illusion is “*something* that is not what it seems to be”. For this reason, Blackmore considers the science of the consciousness as built on false premises. D. Wegner, contrary to S. Blackmore, argues that the experience of conscious will is a feeling that helps us to appreciate our authorship of the things we do. For him, it is important to understand how conscious will might be an illusion and not that it is an illusion as a whole. The conscious self

⁴ A. Damasio (1999) in his theory on the origin of consciousness, self and emotion offers an illuminating evolutionary story of the birth of the autobiographical self from the conscious core self and the unconscious proto-self.

is an illusion in the sense that “the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action” (Wegner 2002, 2). M. Velmans (2004), claims that in spite of Wegner’s insights on how conscious experiences relate to brain processes, it is necessary to outline in what sense conscious will is *not* an illusion. He agrees with Wegner’s causal story of how conscious will arises and enters causally into subsequent mental processing. For Velmans, however, this causal story remains a first-person story. Conscious free will is not an illusion in the sense that this first-person story is compatible with and complementary to a third-person account of voluntary processing in the mind/brain. Finally, the spectrum of ideas eventuates into a need to clarify in what sense an illusion could be considered as *real*.

3 Meeting of minds

T. Metzinger has presented a model of the self based on our everyday (folk psychology) language, which we use when speaking about inner states of ourselves and others.

P. Gärdenfors (2013) presents an account of semantics that is not construed as a mapping of language to the world, but mapping between individual meaning spaces. The meanings of linguistic entities are established via a “meeting of minds.” The concepts in the minds of communicating individuals are modeled as *convex regions* in conceptual spaces. Gärdenfors outlines a mathematical framework based on fixpoints in continuous mappings between conceptual spaces that can be used to model semantics. Considering concepts as a convex will lead, according to him, to a possible agreement for the interactors on a joint meaning even if they start out from different representational spaces. His approach has been contrasted with semantics considered as a mapping between a language and the world or “possible worlds”, which does not involve

the users of the language and says very little about how individual user can “grasp” the meanings determined by such a mapping. Another tradition, cognitive semantics, brings in the language user by focusing on the relations between linguistic expressions and the user’s mental representation of the meanings of the expressions, often in form of “image schemas”. Gärdenfors claims, that cognitive semantics has problems in explaining the social nature of semantics, therefore he proposes a different view of semantics based on a “meeting of minds.” According to this view, the meanings of expressions do not reside in the world or solely in the mental schemes of individual users, but they *emerge* from the communicative interactions between the language users. A meeting of the minds means that the representations in the minds of the communicators will become sufficiently compatible. Gärdenfors assumes that our inner worlds (which surely exist!) can be modeled as spaces with topological and geometric structure. As the main modeling tools he uses *conceptual spaces* (Gärdenfors 2000). The mental space that carries the meanings for a particular individual is partly determined from the individual’s interaction with the world, partly from her interaction with others and partly from her interaction with herself. This view does not entail that different individuals mean the same thing by using an expression, only that their communication is sufficiently successful. Two features of this process should be noted: Firstly, joint attention requires a “theory of mind” that includes second order attention, in the sense that both communicative partners can attend to the attention of the other. Secondly, the two partners need not have the same image of each others’ inner state, because it is possible to reach joint attention.

As an example of how a meeting of minds can be achieved by communication, but without the aid of language, Gärdenfors considers declarative pointing. This act

consists of one individual pointing to an object or spatial location and at the same time checking that the other individual (the “recipient”) focuses his or her attention on the same object or location. The recipient in turn must check that the pointer notices that the recipient attends to the right entity. This attending to each others’ attention is called “joint attention”, it is a mechanism for checking that the minds of the interactors meet in focusing on the same entity. In other words, the *fixpoint*⁵ is reached when one individual sees that the other individual is directed at the correct point and she believes that the other individual is directed to what is pointed at (figure 4).

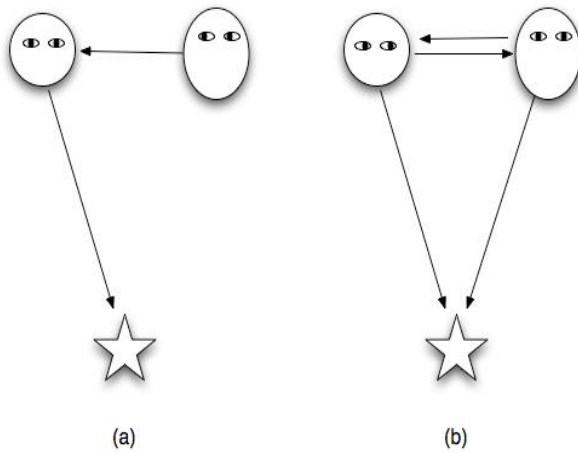


Figure 4: Joint attention as a fixpoint

In sum, speakers may reach a point in which they *feel* there is a mutual understanding.

Concepts considered as closed and bounded regions of Euclidean conceptual spaces, acquire an important topological property of *compactness*. Compactness allows approximating the whole space through a finite number of points. Together with the property of *convexity* Gärdenfors demonstrates an interesting relation to the the prototype theory

⁵ Within „meeting of minds“ semantics fixpoints represent interactive, social process of meaning construction and evaluation.

(Lakoff, 1999). It is illustrated by a *metaphorical composition*.

When concepts are defined as convex regions of a conceptual space, prototype effects are indeed to be expected. In a convex region one can describe positions as being more or less central. In particular, in a Euclidean space one can calculate the centre of gravity of a region. Even if the head and the modifier do not share any dimension, we can still create a mapping between different domains by exploiting the convexity and compactness of the domains.

Indeed, an important implication of the existence of a radial projection is that any two convex compact spaces can be mapped via homeomorphism. This permits the creation of the *metaphor* effect, that is, the transfer of structure from one domain to another. Once this is made, we are back to the previous modifier action. As a simple example, Gärdenfors considers the expression “the peak of a career.” The literal meaning of *peak* refers to a structure in physical space, namely the vertically highest point in a horizontally extended (large) object, typically a mountain. This structure thus presumes two spatial dimensions, one horizontal and one vertical (see figure 5a).

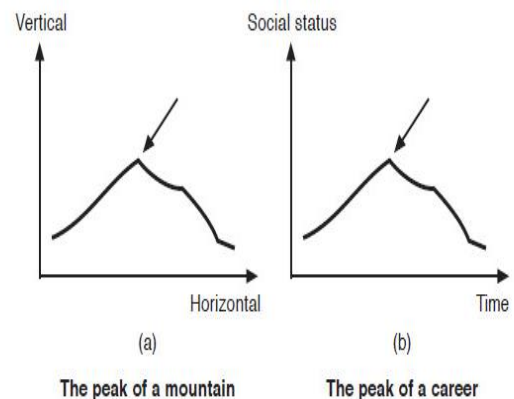


Figure 5: a) A metaphorical reinterpretation of b) the literal meaning of „peak“ (Gärednfors 2000, 177).

A career is an abstract entity without location in space. So how can a career have a *peak*? What happens when we

metaphorically talk about the peak of a career is that the same geometrical structure is applied to a two-dimensional space that consists of the *time* dimension (of the career), which is mapped on the horizontal spatial dimension, and a dimension of *social status* (or “level of accomplishment”, see figure 5b). The latter dimension is normally conceived of as being vertical: we talk about somebody having a “higher” rank, “climbing” in the hierarchy, etc. (Lakoff and Johnson 1980). A metaphor does not come alone – it is not only a comparison between two single concepts but it also involves an identification of the structure of two complete domains. Once a domain has been connected to another via metaphor, this connection may serve as a generator for new metaphors based on the same kind of relations. Metaphorical mappings involve whole systems of concepts. Gärdenfors has presented here a way of composing a modifier and a head which is involved in metaphor and requires a homeomorphic mapping between spaces. The constructions presume that mental spaces are partitioned into domains. In his recent work Gärdenfors concentrates on further and deeper problems within a proposed socio-cognitive type of semantics, evolutionary robotics, structure of events and mutual understanding in communication.

4 Concluding remarks

To what extent diverse philosophical accounts on the nature of the self complement or exclude each other remains a difficult issue with several open questions. Answering them requires clarification of what *can* and what *cannot* be reconciliated among proposed ontological and epistemological perspectives. In the present paper I did not intend to work out a detailed evaluation or criticism of indicated models. My aim was based more on pointing out the importance to acquire transparency between the starting point, used methods and aims in studying

the phenomenon in question. Surely, complexity on the one hand and uniqueness of subjective experience bring a number of difficulties in search for a science of mind. Nevertheless, even at this early stage of studying our inmost intimate inner world it is possible, I think, to hypothesize that:

- it is a necessity to approach states of inner experience as natural phenomena open to scientific methodology and aims of inquiry
- to accept an antireductive stance towards phenomenal experience as a starting point, *before* the very investigation, is a logical and methodological mistake
- improvement of the ways we conceptualize our inner life on an interdisciplinary platform requires a common working characterisation of the phenomenon in question

The strength of outlined models, ideas and theories lies in their explanatory force – by showing a real possibility to model the complex richness of our inner life and *to accept* constraintuitive knowledge about ourselves as highly beneficial.

References

- Blackmore, S.: *Consciousness: An Introduction*. London: Oxford University Press, 2011.
- Blanke, O.: Multisensory brain mechanisms of bodily self consciousness. In *Nature Reviews Neuroscience*, 2012, 13, 556–571.
- Chalmers, D. *The Conscious Mind*. New York: Oxford University Press, 1996.
- Churchland, P. M.: *The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain*. Cambridge: MIT Press, 1995.
- Damasio, A.: *The Feeling of What Happens*. London: William Heinemann, 1999.
- Dennett, D.: *Consciousness Explained*. Boston: Little, Brown, 1991.
- Dretske, F.: *Naturalizing the Mind*. Cambridge, MA: MIT Press, 1995.
- Fauconnier, G., Turner, M.: *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books, 2002.
- Gallagher, S., Zahavi, D.: *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge, 2008.
- Gärdenfors, P.: *Conceptual Spaces: The Geometry of Thought*, MIT Press, Cambridge, MA, 2000.
- Gärdenfors, P.: *How Homo Became Sapiens: On the Evolution of Thinking*, Oxford University Press, Oxford, 2003.
- Gärdenfors, P., Warglien, M.: Semantics, conceptual spaces and the meeting of minds. In: *Synthese*. August 2013, Volume 190, Issue 12, 2165-2193.
- Grady, H., Todd O., Seana C.: Conceptual Blending and Metaphor. In: G. Steen and R. Gibbs (eds.), *Metaphor in Cognitive Linguistics*. Amsterdam/Philadelphia: John Benjamins, 1999.
- Laird, J. P.: *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Harvard University Press, 1983.
- Lakoff, G., Johnson, M.: *Metaphors We Live By*, The University of Chicago Press, Chicago, IL, 1980.
- Lakoff, G., Johnson, M.: *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books, 1999.
- Metzinger, T.: *The Ego Tunnel*. New York: Basic Books, 2010.
- Metzinger, T.: *Conscious Experience*. Thorverton: Imprint Academic & Paderborn, 1995.
- Strawson, G.: Radical Self-Awareness. In: *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*, M. Siderits, E. Thompson, and D. Zahavi (ed.): Oxford: Oxford University Press, 2010, 274–307.
- Varela, F.J.: Neurophenomenology - A Methodological Remedy for the Hard Problem. In: *Journal of consciousness studies*, 1996, 3(4):330-349.
- Velmans, M.: Why conscious free will both is and isn't an illusion. In: *Behavioral and Brain Sciences*, 2004, 27(5): 649-659.
- Velmans, M.: How to define consciousness - and how not to define consciousness. In: *Journal of Consciousness Studies*, 2009, 16(5):139-156.
- Wegner, D.: *The Illusion of Conscious Will*. Cambridge: MIT Press, 2002.

Wittgenstein, L.: *Philosophical Investigations*. Oxford: Blackwell, 1953.

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

Omylnost metakognice o důvodech jednání a její trénink

Ondřej Havlíček, Marcel Brass, Axel Cleeremans, Agnieszka Wykowska

Department of General and Experimental Psychology, LMU Munich
Leopoldstraße 13, 80802 München, Germany
o.havlicek@psy.lmu.de

Department of Experimental Psychology, Ghent University
Henri Dunantlaan 2, 9000 Ghent, Belgium
marcel.brass@ugent.be

CRCN, Faculty of Psychology, Université Libre de Bruxelles
50 avenue F.-D. Roosevelt CP191, B1050 Bruxelles, Belgium
axcleer@ulb.ac.be

Luleå University of Technology
971 87 Luleå, Sweden
agnieszka.wykowska@ltu.se

Abstrakt

Ohledně možnosti a spolehlivosti přístupu k higher-order kognitivním procesům existuje značná skepse. Lidé se například mohou značně mýlit o důvodech svého jednání: V našich experimentech často nedokáží rozlišit, zda byla jejich akce záměrná či automatická. Důvodem nedokonalé metakognice může být i nedostupnost adekvátní zpětné vazby. Právě tu se nám podařilo podat participantům v prezentované studii. Pomocí aplikace fuzzy teorie detekce signálu jsme zjistili, že zlepšili svou metakognitivní citlivost významně více než kontrolní participanti, zatímco rozlišovací kritérium se nezměnilo. Individuální rozdíly v metakognitivních schopnostech tak nemusí být fixní, ale ovlivnitelné. Nacházíme také výrazný rozdíl mezi metakognicí prvního a druhého řádu.

1 Úvod

Metakognici lze charakterizovat jako znalost vlastní mysli. V poslední době značně narůstá zájem o téma metakognice napříč mnoha disciplínami: od invazivní elektrofyziologie u zvířat, přes kognitivní psychologii, po filosofii mysli. Tento zájem lze vysvětlit centrálním významem metakognice pro lidskou kognici, například pro monitorování a řízení vnímání, učení, chování a myšlení obecně, ale možná i sebe-vědomí (Fleming a spol., 2012). Existují značné individuální rozdíly v metakognitivních dovednostech (např., Song a spol., 2011), které lze vztáhnout k rozdílům v neuroanatomii (Fleming a spol., 2010). Není však známo, zda a do

jaké míry je tento anatomický substrát fixní či plastický a s ním spojené metakognitivní dovednosti ovlivnitelné.

Metakognitivních dovedností existuje celá řada a současný výzkum se zaměřuje především na paměťové a percepční, tedy například na schopnost posoudit, nakolik si jsem jistý přesností nějaké vybavené vzpomínky či určitého zrakového vjemu. V našem výzkumu se zaměřujeme na metakognitivní dovednosti více identifikovatelné s tradičním pojetím introspekce, tedy přístup k "higher-order mentálním procesům, jako těm zapojeným v hodnocení, usuzování, řešení problémů a zahajování chování" (Nisbett a Wilson, 1977, s. 232). Nisbett a Wilson jsou, podobně jako mnoho dalších výzkumníků v kognitivní a sociální psychologii, značně skeptičtí k možnosti spolehlivého přístupu k higher-order mentálním procesům. Dle nich máme přístup spíše k jejich produktům (konkrétní soud či chování) a o povaze procesů samotných si tvoříme různé dobré inference, především na základě vlastních apriorních kauzálních teorií o tom, jak by asi takové procesy a různé vlivy fungovat měly. Jedním z důvodů pro tento stav je podle nich nedostatek adekvátní zpětné vazby o fungování naší mysli. Například, vlivy na nějaké jednání mohou být příliš komplexní a nedostupné introspekci a člověk si tak pro sebe může dané jednání post-hoc racionalizovat a de facto konfabulovat o svých motivech, protože z okolního prostředí obvykle nezíská zpětnou vazbu o tom, jaké byly skutečné determinanty tohoto jednání.

V předchozích experimentech (Kühn a Brass, 2009) bylo zjištěno, že lidé v určitých situacích nedokáží rozlišit, zda byla jejich vlastní akce záměrná (deliberovaná) či automatická (reflexivní). V této

schopnosti byly také značně individuální rozdíly mezi participanty. Jedinečnou předností použité úlohy je právě možnost do určité míry "čist mysli" participantů lépe než to dokáží oni sami a určit povahu mentálních procesů vedoucích k jednání. Díky tomu můžeme poskytovat individuální zpětnou vazbu a ověřit tak domněnky Nisbetta a Wilsona o roli zpětné vazby na metakognici a také ověřit míru, do jaké jsou individuální rozdíly minimálně v této metakognitivní schopnosti fixní či ovlivnitelné.

2 Metoda

2.1 Participanti

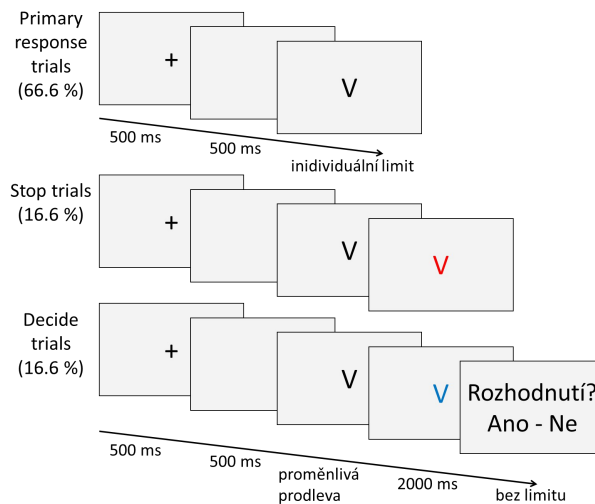
A priori jsme se rozhodli získat data od 40 participantů. Otestovali jsme celkem 51 participantů, z nichž jsme 11 vyloučili na základě apriorních kritérií ověřujících správné porozumění instrukcím a schopnost vykonávat experimentální úlohu. Participanti byli náhodně zařazeni do experimentální či kontrolní skupiny (20 v každé). Věk participantů byl 20 až 33 let ($M = 25,4$), 17 bylo mužského pohlaví, 38 pravorukých. Participanti podepsali informovaný souhlas a za svou participaci obdrželi 8 € za hodinu či zápočet účasti na výzkumu v rámci svého studia. Studie byla schválena místní etickou komisí. Experiment trval celkem 3-4 hodiny.

Participanti byli usazeni v setmělé, tiché místnosti, se sluchátky na uších, 75 cm od LCD monitoru (800×600 pixelů, 60 Hz). Odpovídali pomocí standardní klávesnice. Stimuly byly prezentovány pomocí software E-Prime.

2.2 Procedura

Základní úlohou byla diskriminace zobrazených písmen na obrazovce (cca. 1.3° zřakového úhlu). Na zobrazená písmena "M" a "V" měli participanti co nejrychleji odpovědět levou klávesou a na písmena "N" a "W" pravou. Tato experimentální podmínka představuje tzv. "primary response" (PR), viz Obr. 1. V 1/6 případů se však krátce po zobrazení písmene jeho barva změnila na červenou, signalizující "stop" podmínku. Ta znamená, že participant musí přerušit svou odpověď a nic nestisknout. Prodleva, za jakou došlo k přebarvení písmene, byla řízena schodišťovou procedurou, tak aby úspěšnost zastavení odpovědi byla 50 % (neúspěšné zastavení se označuje jako "failed-to-stop"). Podobně, v další 1/6 případů se barva písmene změnila na modrou, signalizující "decide" podmínku. Participant měl také zastavit odpověď a provést rozhodnutí, zda v tomto konkrétním případě chce odpovědět či ne ("go" vs. "stop"). Následně měl indikovat, zda jeho odpověď (popř. její absence) byla provedena na základě rozhodnutí či pouhé reflexivní reakce. Participanti se měli rozhodnout volně v každé situaci, ale zároveň se rozhodovat přibližně stejně často

pro každou z možností (go nebo stop). Co se počítá za rozhodnutí, bylo detailně vysvětleno (uvědomění barevného stimulu, zastavení odpovědi, rozhodování zda odpovědět či ne, případná reiniciace pohybu). Po skončení experimentu participanti indikovali, že instrukcím velmi dobře rozuměli a bylo jim jasné, co znamenalo se rozhodnout (medián = 6, modus = 7, na stupnici od 1 do 7).



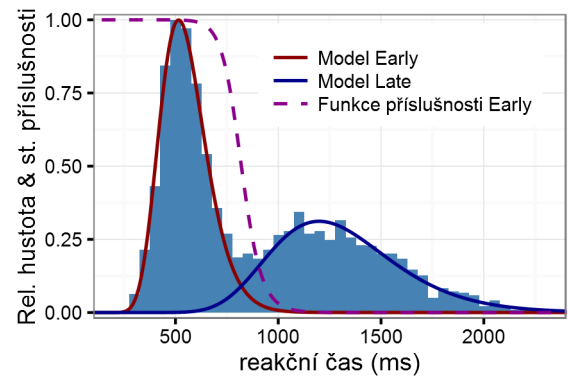
Obr. 1: Struktura trialů a experimentálních podmínek

Experiment probíhal ve dvou po sobě jdoucích dnech. První den participanti provedli 12 bloků po 60 trialech (celkem 720) této úlohy. Tomu předcházely instrukce, procvičení úlohy a měření reakčního času v pouze základní úloze rozlišování písmen. Druhý den participanti obdrželi "metakognitivní trénink", který spočíval v této samé úloze, ale po každém "decide" trialu se na obrazovce objevila zpráva se zpětnou vazbou. V experimentální skupině tato zpráva mohla znít (v angličtině): "Nejspíše jste se rozhodl(a)", "Nejspíše jste se Nerozhodl(a)" nebo "Možná jste se rozhodl(a), možná ne". Volba zprávy se zakládala na reakčním čase v daném trialu. První zpráva se zobrazila, pokud byl reakční čas delší než 0.9627 kvantil rozložení reakčních časů primary response odpovědi předchozí den. Druhá zpráva, pokud byl reakční čas kratší než 0.6049 kvantil a třetí v ostatních případech. Tyto kvantily byly určeny na základě pilotního experimentu ($N=30$) a modelovacích metod popsaných níže, avšak s využitím ostrých klasifikačních prahů tak, aby byla pouze 1% šance mis-klasifikace. Pokud participant v daném trialu neodpověděl, obdržel vždy první zprávu, protože absence odpovědi je v drtivé většině případů participanty označována jako rozhodnutí. Kontrolní skupina obdržela náhodnou z těchto zpráv: "Dobře", "Ne tak dobře", "Ok". Tato zpráva tedy nebyla informativní, ale sloužila jako kontrola emoční valence zpětné vazby a vyvolání nespécifického přemýšlení o úloze. Trénink zahrnoval 6 bloků po 60 trialech (360

celkem). Následoval druhá, zcela stejná část ("session") úlohy jako v první den, tedy opět 12 bloků po 60 trialech. Poté účastníci vyplnili několik dotazníků: Jeden o různých aspektech tohoto experimentu (obtížnost, porozumění, strategie, ...), dále o postojích a přesvědčeních ohledně svobodné vůle (Free will and Determinism Plus, Paulhus a Carey, 2011), locus of control (Internal, Powerful Others, and Chance, Levenson, 1981) a mindfulness (MAAS, Brown a Ryan, 2003).

2.3 Zpracování dat

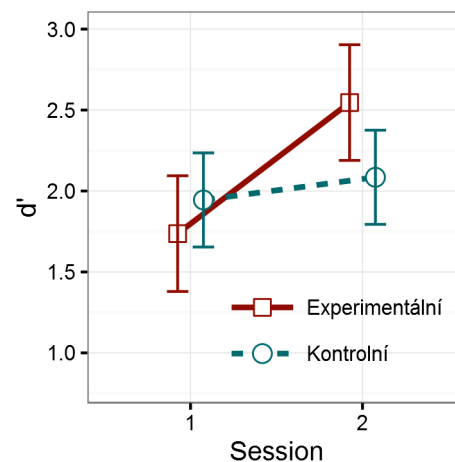
Zajímaly nás případy podmínky "decide" kdy došlo k odpovědi na zobrazené písmeno, a účastníci následně odpověděli, že se rozhodli ("decided-go") nebo že se nerozhodli ("failed-to-decide-go"). Rozložení reakčních časů decided-go trialů vykazovalo jako i v předchozích experimentech jasně bimodální rozdělení, což bylo potvrzeno 1000-krát opakovaným náhodným rozpůlením datového vzorku, maximum likelihood estimation různých modelů (uni- a bi-modální, normální a log-normální) na jedné polovině vzorku, Kolmogorov-Smirnov testy na druhé polovině, a následném srovnání vyhovujících modelů pomocí Akaike Information Criterion, který určil bimodální log-normální rozdělení jako nejlepší vyhovující model dat, viz Obr. 2. První rozdělení ("early decided-go") se přibližně kryje s rozdělením primary response reakčních časů a ostatních případů prostých reflexivních reakcí (failed-to-stop, failed-to-decide-go). Jde tak velmi pravděpodobně o chybné atribuce záměrnosti v podstatě automatickým akcím, viz Kühn a Brass (2009). Pro každého účastníka byla sestrojena klasifikační funkce udávající pravděpodobnost určitého reakčního času jakožto náležitosti do prvního nebo druhého rozložení, tedy jak pravděpodobné je, že byla určitá odpověď provedena reflexivně či na základě deliberace. Takto můžeme probabilisticky klasifikovat decided-go data na zásahy ("late" rozdělení) versus na falešné popluchy ("early" rozdělení) a failed-to-decide-go data na opomenutí (late) versus na správná vyloučení (early). Pomocí aplikace fuzzy teorie detekce signálu (Parasuraman a spol., 2000) tak můžeme spočítat pro každého účastníka dvě metriky: senzitivitu (d' , d' -prime), udávající schopnost rozlišovat mezi záměrnými a reflexivními akcemi, a rozlišovací kritérium (c , criterion), udávající obecnou tendenci označovat akci jako záměrnou či reflexivní.



Obr. 2: Hustota reakčních časů (modely early a late decided-go dat) škálovaná relativně k histogramu modelovaných dat a funkce stupně příslušnosti do early třídy v závislosti na reakčním času.

3 Výsledky

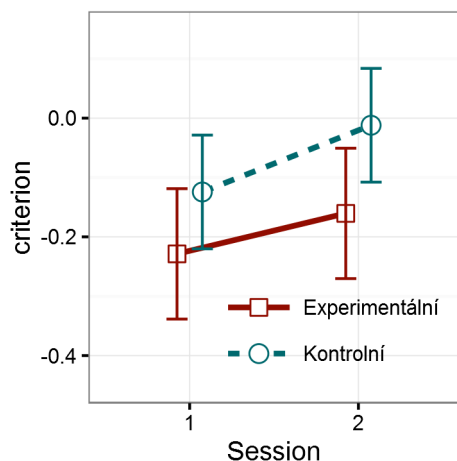
Analýza rozptylu (ANOVA) senzitivity, s within-subjects faktorem "session" (před tréninkem vs. po tréninku) a between-subjects faktorem "group" (experimentální vs. kontrolní skupina) zjistila signifikantní hlavní efekt session, kdy senzitivita byla obecně větší v druhé session ($M = 2,31$, $SD = 1,25$) než v první ($M = 1,84$, $SD = 0,98$; $F[1, 38] = 8,95$, $p = 0,005$), nesignifikantní hlavní efekt skupiny ($F[1, 38] = 0,148$, $p = 0,703$) a signifikantní interakci mezi session a group, $F(1, 38) = 4,49$, $p = 0,0408$. Senzitivita významně vzrostla v experimentální skupině ($t[19] = 3,31$, $p = 0,004$; Hedges' $g = 0,76$, 95% CI [0,25; 1,31]), ale ne v kontrolní skupině ($t[19] = 0,69$, $p = 0,499$; Hedges' $g = 0,11$, 95% CI [-0,21; 0,43]). Změna byla větší v experimentální skupině než v kontrolní skupině (Welch's t-test, $t(36,61) = 2,12$, $p = 0,041$; Hedges' $g = 0,66$, 95% CI [0,03; 1,3]). Metakognitivní trénink tedy specificky zvýšil metakognitivní senzitivitu účastníků. Viz Obr. 3.



Obr. 3: Metakognitivní senzitivita pro obě skupiny účastníků, před a po tréninku (session 1 vs. 2).

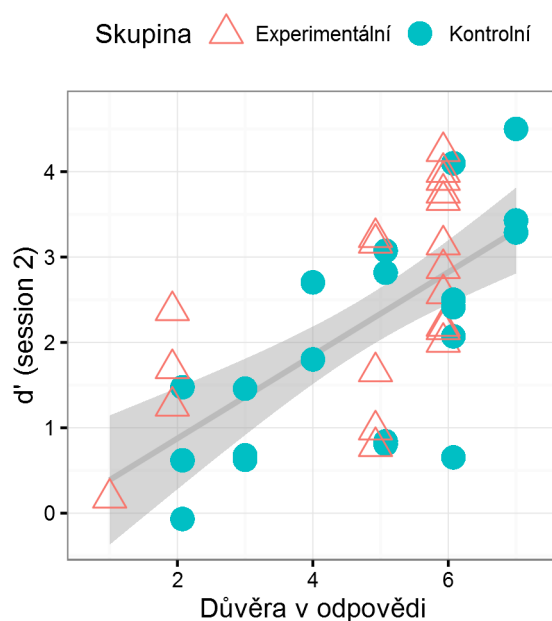
Chybové úsečky znázorňují within-subjects 95% CI pro průměr.

ANOVA rozlišovacího kritéria zjistila signifikantní hlavní efekt session ($F[1, 38] = 3,73, p = 0,0608$), ale nikoli hlavní efekt skupiny ($F[1, 38] = 2,43, p = 0,127$) ani interakci session a skupiny ($F[1, 38] = 0,103, p = 0,750$; Hedges' $g = -0,1, 95\% \text{ CI } [-0,72; 0,52]$), viz Obr. 4. Metakognitivní trénink tedy neměl žádný, či pouze velmi malý vliv na rozlišovací kritérium, obecnou tendenci odpovídat jedním či druhým způsobem.



Obr. 4: Rozlišovací kritérium pro obě skupiny participantů, před a po tréninku (session 1 vs. 2). Chybové úsečky znázorňují within-subjects 95% CI pro průměr.

Analýza dotazníků je mimo záběr tohoto příspěvku, mimo jedné otázky týkající se vztahu mezi skutečnou a domnělou metakognitivní citlivostí, tedy jakési "meta-metakognice" či metakognice "druhého řádu". Po skončení experimentu zodpověděli participanté otázku, jak velkou důvěru měli obecně ve své odpovědi na otázky ohledně záměrnosti svých rozhodnutí. Překvapivě, ačkoli participanté vykazovali různě dobrou či špatnou senzitivitu, jejich hodnocení jistoty s ní značně korelovalo. Spearmanova korelace udávané jistoty s metakognitivní senzitivitou v druhé session (čerstvější zkušenost k posouzení): $\rho = 0,647, 95\% \text{ CI } [0,420; 0,798], p < 0,001$; v první session: $\rho = 0,536, 95\% \text{ CI } [0,269; 0,753], p < 0,001$. To znamená, že metakognice druhého řádu většiny participantů byla poměrně vysoká: dobře znali, nakolik znali své mentální procesy. Několik participantů se přečeňovalo a pouze malé množství podčeňovalo, viz Obr. 5.



Obr. 5: Vztah skutečné senzitivity (d') v druhé session experimentu a posuzované důvěry ve svou metakognitivní rozlišovací schopnost. Linie zobrazuje lineární model, stínovaná oblast 95% CI.

4 Diskuse

Cílem studie bylo zjistit vliv zpětnovazebního učení na metakognitivní dovednosti participantů, konkrétně v úloze, ve které se mnoho participantů dopouští chybné atribuce důvodů svého jednání. Studie úspěšně replikovala výsledek, že mnoho lidí může označovat zcela automatickou, reflexivní odpověď za záměrnou, provedenou po vědomé registraci signálu k rozhodnutí, inhibici akce a deliberaci o její reiniciaci. Ani detailní instrukce a exkluze participantů, kteří je dostatečně nedodržovali (např. poměr rozhodnutí pro go vs. stop v určitém pásmu kolem 50:50), tento výrazný efekt neodstranily.

Novým aspektem této studie byla individuální, probabilistická klasifikace jednotlivých rozhodovacích trialů, na místo skupinové, ostré klasifikace v předchozím výzkumu. Mohli jsme tak zohlednit jak nejistotu v inferenci z reakčních časů, zda byla daná odpověď výsledkem pouze reflexivních nebo deliberačních procesů, i vágnost (fuzziness) samotného konceptu rozhodnutí, které může trvat dlouhou, ale i poměrně krátkou dobu. Také jsme díky tomu mohli využít formalizmu teorie detekce signálu a rozlišit, zda jsou misatribuce dány nízkou metakognitivní senzitivitou či pouhou individuální tendencí k určitému typu odpovědi.

Hlavním cílem studie bylo ověřit, zda podávání zpětné vazby o povaze mentálních procesů stojících za akcí dokáže ovlivnit metakognitivní citlivost lidí. Naše

výsledky svědčí o kladné odpovědi. Minimálně námi zkoumané metakognitivní dovednosti tak nejsou dány fixními individuálními odlišnostmi v neuroanatomii, ale jsou plastické a ovlivnitelné. Zlepšení však nebylo absolutní a i po tréninku docházelo k relativně velkému množství atribucí záměrnosti reflexivním akcím.

Je nutno vzít v potaz, že kvůli časové náročnosti úlohy jsme získali vzorek pouze 40 lidí, což nám umožnilo detekovat specifický efekt metakognitivního tréninku se statistickou silou "pouze" 0,54 a odhad skutečné velikosti efektu je tak poměrně široký. Je třeba budoucího výzkumu k ověření a zpřesnění našich výsledků. Také je třeba vzít v potaz ekologickou validitu naší studie. Úloha byla cíleně vytvořena tak, že v ní může k chybné atribuci snadno docházet. Většina reálných situací takový charakter nemá. Kognitivně-psychologický výzkum využívá takovýchto experimentů jako jakési lupy na odhalování charakteru kognitivních (a v tomto případě metakognitivních) procesů a souvisejících vlivů. Aplikace těchto základních poznatků je pak také otázka pro budoucí výzkum.

Zajímavým incidentálním zjištěním studie byla poměrně dobrá meta-metakognice u většiny participantů, včetně těch s nízkou metakognicí v dané úloze. To může naznačovat, že existuje více různých, na sobě do jisté míry nezávislých metakognitivních procesů monitorujících fungování naší mysli a našeho jednání.

Poděkování

Tento příspěvek vznikl za podpory Graduate School of Systemic Neurosciences, LMU Munich pro OH a za podpory German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), individuální grant (WY-122/1-1) pro AW.

Literatura

Brown, K. W. a Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4): 822–848.

Fleming, S. M., Dolan, R. J. a Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594): 1280–1286.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. a Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998): 1541–3.

Levenson, H. (1981). Differentiating among internality, powerful others, and chance. Ve sborníku *Research*

with the Locus of Control Vol. 1., str. 15–63. New York: Academic Press.

Kühn, S. a Brass, M. (2009). Retrospective construction of the judgement of free choice. *Consciousness and Cognition*, 18(1): 12–21.

Nisbett, R. E. a Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3): 231–259.

Parasuraman, R., Masalonis, A. J. a Hancock, P. A. (2000). Fuzzy Signal Detection Theory : Basic Postulates and Formulas for Analyzing Human and Machine Performance. *Human Factors*, 42(4): 636–659.

Paulhus, D. L. a Carey, J. M. (2011). The FAD-Plus: measuring lay beliefs regarding free will and related constructs. *Journal of Personality Assessment*, 93(1): 96–104.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S. a Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition*, 20(4): 1787–1792.

The encoding of proprioceptive inputs in the brain: knowns and unknowns from a robotic perspective

Matěj Hoffmann

Istituto Italiano di Tecnologia
Via Morego 30
16163 Genova, Italy
Email: matej.hoffmann@iit.it

Nad'a Bednářová

Fakulta Elektrotechnická, ČVUT v Praze
Žitkova 1903/4
166 36 Praha
Email: nada.bednarova@gmail.com

Abstract

Somatosensory inputs can be grossly divided into tactile (or cutaneous) and proprioceptive – the former conveying information about skin stimulation, the latter about limb position and movement. The principal proprioceptors are constituted by muscle spindles, which deliver information about muscle length and speed. In primates, this information is relayed to the primary somatosensory cortex and eventually the posterior parietal cortex, where integrated information about body posture (postural schema) is presumably available. However, coming from robotics and seeking a biologically motivated model that could be used in a humanoid robot, we faced a number of difficulties. First, it is not clear what neurons in the ascending pathway and primary somatosensory cortex code. To an engineer, joint angles would seem the most useful variables. However, the lengths of individual muscles have nonlinear relationships with the angles at joints. Kim et al. (Neuron, 2015) found different types of proprioceptive neurons in the primary somatosensory cortex – sensitive to movement of single or multiple joints or to static postures. Second, there are indications that the somatotopic arrangement (“the homunculus”) of these brain areas is to a significant extent learned. However, the mechanisms behind this developmental process are unclear. We will report first results from modeling of this process using data obtained from body babbling in the iCub humanoid robot and feeding them into a Self-Organizing Map (SOM). Our results reveal that the SOM algorithm is only suited to develop receptive fields of the posture-selective type. Furthermore, the SOM algorithm has intrinsic difficulties when combined with population code on its input and in particular with nonlinear tuning curves (sigmoids or Gaussians).

1 Introduction

Although proprioception does not belong to the commonly held “five senses”(sight, hearing, smell, touch, taste) view that dates back to Aristotle, it is, along with sense of balance for example, of utmost importance for our everyday existence. What is proprioception actu-

ally? The term has its origin in the work of Sherrington (1906) who stated: “In muscular receptivity we see the body itself acting as a stimulus to its own receptors—the proprioceptors”. In short, we will equate it with *sense of limb position and movement* (see (Proske and Gandevia, 2012) for an extensive survey). The reason why proprioceptive sensations are receiving so little attention in our everyday experience is mainly because we are largely unaware of them, which in turn can be attributed to their predictability. Unlike “exteroceptors”, such as sight or hearing that focus on events in the external environment that are not fully under our control, the position and movement of our limbs is something that is typically under our control and thus, these sensations typically receive less (conscious) attention (Proske and Gandevia, 2012). A related term is *kinesthesia*, which strictly means movement sense, but has been sometimes equated with proprioception.

2 Representation of proprioception in biology

2.1 Receptors

The key receptors in proprioception are muscle and skeletal mechanoreceptors: muscle spindles, Golgi tendon organs, and joint capsule mechanoreceptors—an overview is provided in Table 1. They sense different physical quantities: while the muscle spindles and joint capsule mechanoreceptors are sensing kinematic variables (position and velocity), the Golgi tendons are sensing dynamic (or kinetic) quantities (forces). In addition, stretch-sensitive receptors in the skin (Ruffini endings, Merkel cells in hairy skin, and field receptors) also signal postural information Kandel et al. (2000, p. 443).¹ However, in summary, “the modern view has muscle spindles as the principal proprioceptors”(Proske and Gandevia, 2012).

¹Muscle stretch can roughly be equated with muscle length.

| Receptor | Fiber group / name | Submodality |
|--------------------------------|--------------------|--|
| Muscle spindle primary | $A\alpha/ Ia$ | Muscle length and speed |
| Muscle spindle secondary | $A\beta/ II$ | Muscle stretch (\sim length) |
| Golgi tendon organ | $A\alpha/ Ib$ | Muscle contraction / tension (\sim force) |
| Joint capsule mechanoreceptors | $A\beta/ II$ | Joint angle |

Tab. 1: Muscle and skeletal mechanoreceptors.

2.1.1 Muscle spindles as principal proprioceptors

Muscle spindles respond in terms of their firing rate to changes of muscle length (that is the dynamic response – muscle spindle primary) as well to a sustained length of the muscle (static/tonic response – muscle spindle primary and secondary). “Position sense can therefore be envisaged as signaled by the mean rate of background discharge in muscle spindles, including that generated by both primary and secondary endings” (Proske and Gandevia, 2012). However, the position signal delivered by individual spindles seems to be far from perfect. For example: (i) muscle spindles have also efferent innervation (fusimotor – gamma neuron activity) and thus there are differences between active and passive muscle stretching, (ii) some spindles generate background activity at all muscle lengths, no matter how short the muscle; others fall silent at short lengths. It is also not clear how the limits are retrieved and absolute position information seems to be inaccurate. Such effects may be mitigated if population of spindles are considered and possibly also by recruiting the information from joint receptors (acting possibly as limit detectors). The interested reader is referred to (Jones et al., 2001; Macefield, 2005; Proske and Gandevia, 2012).

2.1.2 From muscle length to joint angle

The spindles deliver information about the length of a single muscle – flexor or extensor – acting over a particular joint (or often also two joints in case of biarticular muscles). Not surprisingly, every spindle afferent is directionally tuned to movement of the joint in a particular direction—as can be inferred from the arrangement of its parent muscle (Jones et al., 2001). However, in order to get information about the joint angle, one has to also consider that there is a nonlinear relationship between the muscle lengths and the angle at the joint. This is illustrated in Fig. 1 for the elbow joint (after (Shadmehr and Wise, 2005)). The length of triceps, λ , is linked to θ , the supplementary angle to the elbow angle ϕ , by the following equation:

$$\lambda = \sqrt{a^2 + b^2 - 2 \cdot a \cdot b \cdot \cos(\theta)} \quad (1)$$

Where and how (or if) the inverse of this transformation (to acquire θ and eventually ϕ from λ) is performed by spinal or cortical neurons has to our knowledge not been described in the literature.

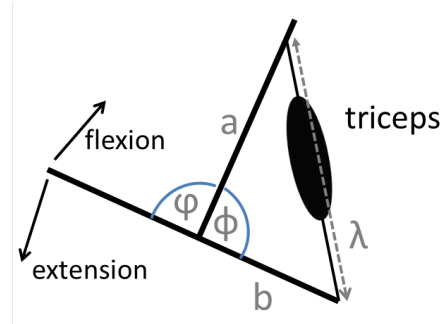


Figure 1: Geometry of elbow joint and triceps muscle. After (Shadmehr and Wise, 2005).

To encode the 3D configuration of a complex joint (such as wrist or ankle), taking into account signals from a population of spindles of different muscles acting over the joint seems indispensable (Jones et al., 2001; Bergenheim et al., 2000)—but again, it is unclear how and where would this information be combined to give an explicit representation of such 3D joint configuration.

2.2 Neural encoding

As described above, there seems to be consensus that muscle spindles are the principal proprioceptors. However, it still remains unclear what the encoded variables are (in particular whether they are the lengths of individual muscles rather than joint angles). Second, we have also not found in the literature which neural coding type is applied to the proprioceptive modality. We will focus on rate codes—where mean neural firing rate carries information—rather than temporal (spiking) codes. The general rate types are:

- *Intensity code.* The activity of the neuron is proportional to (or a monotonic function of) the coded variable (Mallot, 2013).
- *Local coding scheme.* The sensory space is divided into nonoverlapping areas that can be resolved by small groups of topographically arranged neurons with small receptive fields (Ghazanfar et al., 2000). Mallot (2013) calls this *channel coding* or *labeled line code without overlap*.
- *Distributed (population) coding scheme.* Neurons have relatively large and overlapping receptive

fields compared with the sensory resolution measured behaviorally (Ghazanfar et al., 2000). (Malot, 2013) calls this *labeled line code with overlap* and the “standard way” of neural coding.

The encoding type is also intertwined with different *neural tuning curves* that relate the average firing rate of a neuron to relevant stimulus parameters. The tuning curves cited in the literature are:

- *Linear*
- *Sigmoidal*
- *Gaussian*

All of the above will be implemented and illustrated later in this article in Section 4.2. All tuning curves have specific properties which may be more suited to represent a stimulus with particular characteristics. In the linear case, firing rates are a linear function of the stimulus. Different neurons would have different equations (in terms of slope and y-intercept) (see e.g., (Zipser and Andersen, 1988)). Sigmoidal tuning curves are also monotonic but nonlinear functions of the input. Gaussian tuning curves feature a peak—are nonmonotonic functions—and may best encode the stimulus at this tuning curve peak. This may be useful if the input is noisy, since high firing rates are the neuron’s most distinct response. Alternatively, stimulus may be coded not at the peak but at the high-slope regions of the tuning curve where the discrimination is the highest (see (Butts and Goldman, 2006) for an information theoretic analysis). The latter applies both to Gaussian and sigmoidal curves (not the linear functions where the slope is constant). However, we have not found in the literature what the tuning curves used for proprioceptive stimuli are.

2.3 Ascending neural pathway

The principal ascending neural pathway from the receptor sites to the cortex is common with that carrying “fine touch” and it is the so-called medial lemniscal pathway with “relay stations” at dorsal root ganglion, medulla, and thalamus, before reaching the neocortex, mostly terminating in area 3a.

2.4 Representation in the cortex

Before we briefly review the findings regarding the representation of proprioception in the cortex, it is worth noting that the findings about neuronal responses and receptive fields obtained using electrode recordings in monkeys often come from rather unspecific stimulation of so-called deep receptors: “light to moderate taps, digit and limb manipulation, and light pressure, categorized as ‘deep stimulation’, were used to stimulate the muscles, joints, and skin.”(Seelke et al., 2011) A notable exception is (Kim et al., 2015) who have used precisely controlled proprioceptive and cutaneous (or

“tactile”) stimulation. They also note that unlike cutaneous sensory processing, cortical mechanisms underlying proprioception are poorly understood.

2.4.1 Anterior parietal cortex

The anterior parietal cortex encompasses areas 3a, 3b, 1, and 2 and is also commonly referred to as primary somatosensory cortex (SI), even if that may be ambiguous (some consider only area 3b as primary). From these areas, area 3a is considered the most primary receptive site for proprioceptive inputs (while area 3b is the “most primary” cutaneous/tactile). “Area 3a ultimately receives input from group Ia muscle spindle afferents, and contains neurons that respond to the stimulation of these and other deep receptors in the skin.” (Huffman and Krubitzer, 2001). (Kim et al., 2015) precisely controlling movements of individual digits (fingers) in a macaque monkey have identified three types of proprioceptive neurons:

- *Single-digit position scaled*
- *Multi-digit position scaled*
- *Posture-selective*

The first type, single-digit position scaled, means that the neuron responds in a monotonic fashion to the position (i.e. angle in a particular axis) of a single digit. This would probably be most compatible with the intensity code and a linear tuning curve. The second type, multi-digit position scaled, is sensitive to positions of multiple digits, which is still suggestive of a rate code, but some additive mechanism for multiple joints is necessary. The posture-selective neurons, on the other hand, do not scale across specific a dimension. They are responsive to a particular posture—a specific conformation of the hand, i.e. specific position of a number of different joints. Such neurons would be compatible with either local or population coding scheme and perhaps go most naturally with Gaussian tuning curves (the peaks of tuning curves at specific positions of every joint). (Kim et al., 2015) found all of these neuron types throughout all of SI: they were most abundant in area 3a, but present also in 2, 1, and 3b. While their results were obtained for the hand receptive fields only, a similar pattern may perhaps be expected also for other body parts—arm joints, for example.

Regarding topography, (Krubitzer et al., 2004) using a macaque monkey found that area 3a contains a complete representation of deep receptors and musculature of the contralateral body and that the general organization of body part representations mirrors that of area 3b. However, the topographic organization is less precise and less orderly in 3a and some representations, such as the hand, are split or fractured like maps of motor cortex.

2.4.2 Posterior parietal areas

Whereas in the anterior parietal cortex, the receptive fields of neurons typically span individual joints or combinations of a few adjacent joints, in the posterior parietal areas, the receptive fields are in general larger. Body posture is primarily encoded in the superior parietal lobule (areas 5 and 7), a putative homologue of the so-called parietal reach region in the monkey Pellijeff et al. (2006). Seelke et al. (2011), for example, analyzed the receptive field (RF) of neurons in posterior parietal area they termed 5L (lateral) of the macaque and found representation of “deep receptors” of the limb. The RFs of individual neurons often spanned multiple joints, such as elbow and wrist or shoulder and elbow.

2.5 Development in the fetus and newborn

Somatosensation and its representation starts developing already in the womb. In fact, the somatosensory modality starts functioning over the whole body from as early as 17th gestational age and before other sensory modalities Bradley and Mistretta (1975), although connections to the neocortex are not present until the 20th gestational week. Please refer to Milh et al. (2007) and references therein for a more detailed account. Spontaneous motor activity is playing an important role. From about 22 weeks, touching the mouth is already different from touching eyes (more delicate). However, Krubitzer et al. (2004) using macaque reported that unlike 3b (~ “primary tactile”) in which the topography seems to be present in neonates, area 3a (and 1 and 2) are not responsive in neonates in New and Old World monkeys. Furthermore, and again unlike 3b, the construction of area 3a is to a large extent based on the use of a particular body part rather than innervation density (as a matter of fact, about half of the muscle spindles in the body are located in the neck (Cole and Paillard, 1995)).

3 Computational and robotic modeling of proprioception

There has been some computational and robotic modeling addressing the development of somatosensory representations, even though more work has focused on the tactile rather than proprioceptive modality. The work of Kuniyoshi, Mori, and colleagues has addressed prenatal development in foetus simulator including a uterine environment. Their models can also be said to be the most biologically motivated. The foetus simulator has 198 muscles and is actuated with a central pattern generator model. It is equipped with 764 tactile sensors as well as muscle spindles and Golgi tendon organs. Different works addressed the development of SI (Yamada et al., 2013) or a hierarchy of SI and SII using a denoising autoencoder (Sasaki et al., 2013). Aflalo and

Graziano (2006) investigated motor cortex development using a Self-Organizing Map. Drix and Hafner (2014) studied learning proprioceptive and motor features from joint angle values in a quadruped robot using a neural “intrinsic plasticity” rule. Finally, Žabkar and Leonardi (2016) presented a learning framework that builds compositional hierarchies of the joint angle space of the Nao robot.

4 Modeling on iCub humanoid robot and Self-Organizing Maps

We have studied the development of proprioceptive representations using the simulator of the iCub humanoid robot (Metta et al., 2010; Tikhonoff et al., 2008). Here we present an extension of the results of Bednářová (2015).

4.1 Data collection

Data was collected during a body babbling phase. However, we did not employ random body/motor babbling, which would not induce any structure in the proprioceptive space (space of joint angles). Instead, we were inspired by Aflalo and Graziano (2006) who, studying the possible motor cortex development, employed so-called ethological categories: coordinated movements that the monkey uses often (e.g., hand-to-mouth, manipulation in central space in front of the monkey). Thus, we have created a simple data set where the iCub moves the hand in a region of 20x15x15 cm in front of its face and follows it with its gaze (foveating with both eyes). That is, the positions were chosen randomly in the Cartesian space and then the iCub Cartesian and gaze controllers (Pattacini, 2011) were employed to move the hand to the position and follow it by gaze. Thus, joint positions were not randomly chosen, but corresponded to coordinated behavior inspired by a typical “ethological category” of an infant. Joints were sampled at 50 Hz and there were 7 arm joints (3 shoulder, 2 elbow, 2 wrist) of the left arm and 6 joints of the neck and eyes plant (neck pitch, roll, yaw; eyes tilt, version, vergence), giving 10 DoF (Degrees of Freedom) in total. The data set employed had a 20 min. duration.

4.2 Encoding iCub joints data

Before feeding the data to a Self-Organizing Map (SOM), we have implemented all three types of neural encoding known from biological neural systems, as reviewed in Section 2.2. In addition, we have also studied the case where the joint values are only normalized (taking the range of individual joints into account) and fed directly into the SOM.

In order to determine the number of tuning curves to encode every DoF (every DoF having a diffe-

rent range), there are two options: (a) set the number of tuning curves constant, or (b) set the offset between adjacent tuning curves constant. In both cases, the other variable can then be calculated correspondingly. In what follows, we have experimented with both variants. In all expressions below, x represents the original angle values in degrees and y represents the encoded value. The output range is in the interval $\langle 0, 1 \rangle$.

4.2.1 Linear functions

Individual tuning curves were simply given by the equation

$$y = (a \cdot x) + b$$

with varying parameters for the slope a and y-intercept b . Fig.2 illustrates the whole population coding for a single DoF with 20 tuning curves: 10 with positive and 10 with negative slopes.

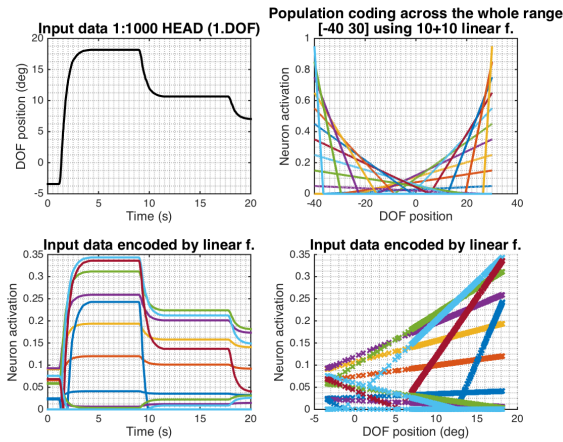


Figure 2: Movement of one DoF encoded by linear tuning curves. (Top left) Input data. Joint position over a 20 s interval. (Top right) Tuning curve activations on input space. (Bottom left) Input data encoded by individual tuning curves. (Bottom right) Close-up of part of the input space and its encoding.

4.2.2 Sigmoids

A sigmoidal tuning curve is given by the following equation:

$$y = \frac{1}{1 + e^{sgn \cdot (-x + offset)}}$$

where sgn sets the curve orientation (positive or negative), $offset$ represents the position of the inflexion point and x are the input values.

As noted above, there are two options for setting up tuning curve parameters for the population. If we set up a fixed number of curves encoding each DoF, the offsets for every DoF for sigmoids in one direction are cal-

culated as:

$$offset = \frac{(max_range - min_range)}{(number_of_sigmoids)}$$

An illustration of the encoding for a single DoF is depicted in Fig. 3.

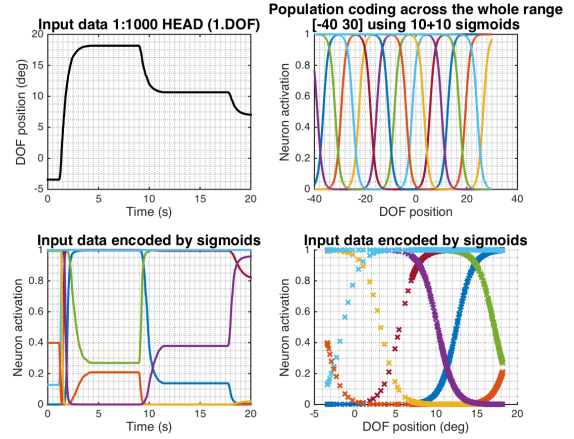


Figure 3: Movement of one DoF encoded by sigmoidal tuning curves. See Fig. 2 for a description of all panels.

4.2.3 Gaussian curves

Gaussian tuning curves are described by the following equation:

$$y = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

with the curve's center or peak at μ , σ the distance to the inflexion point and x the input.

In this case, the curves are symmetrical about the center and we will need only one set of tuning curves (no positive and negative). The offsets between curves can be determined similarly to the previous section. The additional parameters, μ and σ , for the population will then be calculated as follows:

$$\sigma = \frac{(max_range - min_range)}{(number_of_gauss - 1)}$$

$$\mu = min_range + \sigma + offset$$

An illustration of the encoding for a single DoF is depicted in Fig. 4.

4.3 Decoding

Although the inverse mapping to the encoding may not be strictly required in all cases when stimuli are encoded by the brain, it seems that it may often be useful. In our particular application where the data are further processed by a SOM algorithm, decoding is necessary to visualize and quantify what the network has learned. Both

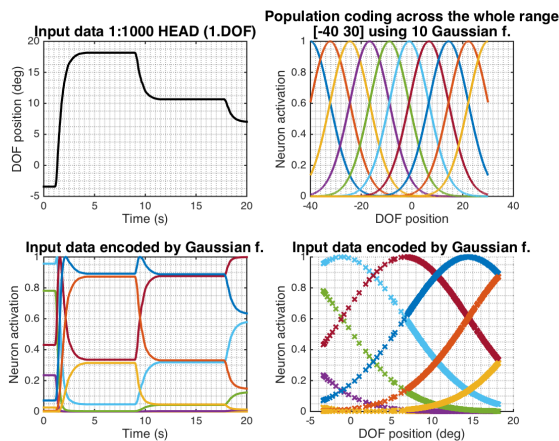


Figure 4: Movement of one DoF encoded by Gaussian tuning curves. See Fig. 2 for a description of all panels.

the linear and sigmoidal tuning curves are monotonic functions and thus a proper inverse exists. Conversely, the Gaussian tuning curve is a nonmonotonic, many-to-one function, that does not have a proper inverse. The formulas for obtaining the inverse are specified below.

4.3.1 Linear functions – decoding

$$x = \frac{(y - b)}{a}$$

4.3.2 Sigmoids – decoding

$$x = -\frac{\ln(1 - y) - \ln(y) - offset \cdot sgn}{sgn}$$

4.3.3 Gaussian curves – decoding

$$x = \mu + \sqrt{-\ln(y) \cdot 2\sigma^2}$$

4.4 Self-organizing map on proprioceptive data

A Self-Organizing Map (SOM) (also self-organizing feature map or Kohonen map) algorithm (Kohonen, 1982, 1990) was used to process the encoded or normalized only joint angle data. A freely available SOM toolbox (Vesanto et al., 1999) was used.

4.4.1 Problems with population coding as SOM input

Let us consider one DoF with movement range from -40 to 30 degrees. Firstly, we encode the movement of this joint using 10 Gaussian curves, then we process the encoded values by a SOM, and finally we use backward transformation for decoding the result back to angles.

Let us also consider a simplified version of SOM with only one output neuron. The input space after encoding has thus 10 dimensions and there are also 10 weight vectors from the input space to the single output neuron. The range of the input variables as well as weight vector components is $< 0, 1 >$. Standard SOM random initialisation takes all input data, calculates the range of every input dimension in the dataset and randomly picks one number from that range to initialize the weights. A property of the SOM algorithm is that the weight vector of an output neuron can be interpreted as a representative point in the input space. However, recall that the 10 dimensions are a population vector derived from a single DoF value. Hence, initializing every weight vector randomly will have the consequence that this point in the input space does not correspond to any angle of the joint, as illustrated in the middle panel of Fig. 5 for the case of Gaussian curves. Therefore the initialisation of SOM had to be modified to first randomly pick an angle from the joint angle range, encode it using the respective encoding and use these values to seed the SOM weight vector—see Fig. 5 right panel.

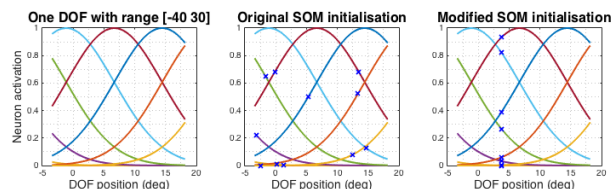


Figure 5: Modified SOM initialisation. See text for details.

However, this modification does not warrant that the representation will be consistent—in the sense that the representative vectors of the SOM output layer will correspond to joint angles—after learning. During training, the SOM algorithm iterates over vectors from the input space and everytime determines the most similar weight vector among the output layer neurons. Such neuron becomes the so-called best matching unit (BMU) which updates its weights in order to decrease the distance to the input presented (neurons in the BMU’s neighborhood are also updated). This is illustrated in Fig.6 using Gaussian curves. The left panel depicts an input vector and its encoding. The middle panel a hypothetical initialisation of the SOM weight vector to a different angle. Finally, the right panel depicts the situation after a training step – weight vector update. The weights are drawn toward the input vector in the encoded space. However, since the different tuning curves have different slopes at the different activation values, the updated vector will not correspond to any real joint angle on the input, as illustrated by the green marks in the right panel. With Gaussian curves, there is the additional complication that the inverse mapping is not unique. This is not the case for sigmoids and linear tuning curves. However, the problem is still present, even for the li-

near case: even if the slope of a single curve is constant, since the pool contains lines with different slopes, the reverse transformation will also not give a unique angle of the DoF under consideration. This situation is likely to worsen with every training step.

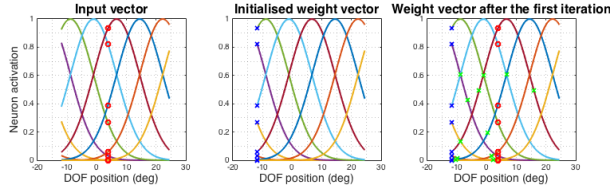


Figure 6: Inconsistency of population coding and SOM weight vector update.

In this case, if one wants to decode from the SOM’s weight vector to the original space of joint angles, that is decode the values of a population, there is no straightforward solution. The method that worked the best for us was non-parametric Kernel Density Estimation² (KDE) that estimates the probability distribution of input data using the following expression:

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right); -\infty < x < \infty$$

where h represents so called ‘bandwidth’ – width of the window, influencing the ‘smoothness’ of the curve, K_h represents normalised probability function, for example Gaussian, and $x_i, i = 1..n$ is the input data. An example is provided in Fig. 7.

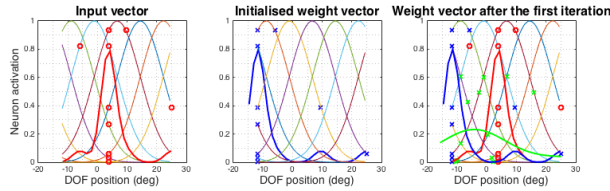


Figure 7: Kernel density estimation for decoding the population code from Fig. 6.

4.4.2 Training a 5x5 SOM from the 10 DoF data set

Using the data set specified in Section 4.1, we trained a SOM with a 5x5 output lattice. Various combinations of encoding (linear, sigmoid, Gaussian curves, or normalized inputs only), order of inputs as presented during training (ordered as in data set or randomly shuffled), and training length (number of reiterations of the whole input dataset in one map training) were tested. The connections from the input layer to the SOM were all-to-all, that is, every output neuron of the map was learning to

²<http://www.mathworks.com/help/stats/kernel-distribution.html>

represent a point in the space of all DoF—a complete posture of the robot in terms of arm and head.

‘Random’ input data order as well as longer training length always gave us better results, so we will report results for this setting only. We also tested the two variants of population coding setup—fixed number of curves per DoF or fixed offset between curves. Tests showed that the resulting map has better quality if we set up constant number of functions encoding every DoF. This guarantees that every DoF will have an equal number of input neurons to the SOM.

Various measures have been proposed to numerically assess the organization of trained SOMs (for an overview, see (Polani, 2002) and references therein). In general, they either assess the vector quantization or the topology preservation capability of the SOM algorithm. We have experimented with different measures. First, Fig. 8 provides a qualitative insight by visualizing one of the best learned maps (with Gaussian tuning curves). The representative vectors of every output neuron were decoded back to the joint angle space and visualized on the iCub simulator. Some topology preservation in the posture space is apparent.



Figure 8: Visualization of SOM after training and decoding to angles. Every cell of the 5x5 matrix illustrates the posture represented by each of the output layer neurons. The decoded DoF positions were visualized on the iCub simulator.

A quantitative insight is provided in Fig. 9. Quantization error measures the quality of input data representation by calculating the average distance of every input vector with its BMU vector from the learned map. We have evaluated the criterion for different types of population coding and different number of tuning curves. In addition, we tested the case in which the joint angle values are only normalized and passed directly to the SOM. Note that in this case, the input space has only 10 dimensions (10 DoF), whereas in the other cases where

population coding is applied, the input space dimension is $10 \times nr.$ curves encoding a single DoF.

The results clearly demonstrate that all population encoding variants lead to worse results compared to the case where normalized inputs were directly fed to the SOM. This may be surprising at first, but it is a consequence of the problems illustrated in Section 4.4.1. The problem is amplified if more tuning curves per DoF are used. Furthermore, the situation gets also worse for the nonlinear tuning curves (sigmoids and Gaussians). This confirms that there is an inherent incompatibility between population coding and the adaptation step of the SOM algorithm.

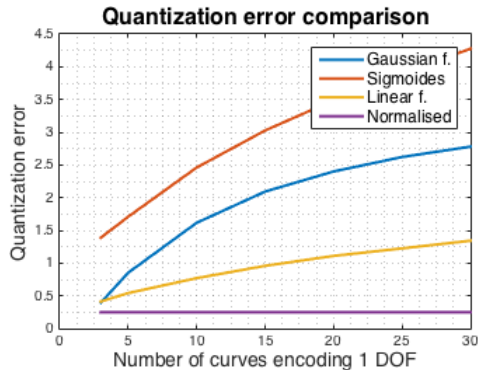


Figure 9: Comparison of quality of learned maps in terms of quantization error. Results for different encoding types as well as different number of tuning curves are plotted. For the sigmoids and linear functions, the nr. curves per DOF corresponds to their count with one orientation—the total number was then double. Random input order and training length of 6 cycles was used in all cases.

5 Discussion, open questions, future work

This article without doubt raises more questions than provides answers. With the the goal of providing a computational model of the representation of proprioception in the brain and its development, it falls significantly short of its aspirations. However, we feel that a number of important issues have been identified, which we discuss below.

First of all, empirical knowledge regarding the workings and development of proprioception and its representation in biology is fragmented and partially contradictory. Or, perhaps, such is the information that the the neural systems dealing with proprioception are processing. Feldman (2009), for example, concludes that “ambiguity of positional information is characteristic of all afferent signals involved in kinesthesia“. We want to make a number of observations. First, the view of isolated bottom-up processing of proprioceptive information is wrong. There is now sufficient evidence sug-

gesting that processing somatosensory information is a “team effort”(Saal and Bensmaia, 2014), where tactile and proprioceptive (Kim et al., 2015) and even vestibular and downstream (top-down) information (Huffman and Krubitzer, 2001) is integrated. Area 3a—even if the most “proprioceptive“ and most primary—can thus hardly be considered as an isolated area processing ascending proprioceptive inputs. Second, the mechanism of formation or development of the proprioceptive representations is unclear. As pointed out earlier, there is evidence suggesting that the topography of area 3a in monkeys is to a large extent learned after birth (Krubitzer et al., 2004). However, once learned, its topography is very similar to that of neighboring area 3b receiving primary tactile inputs. This may seem surprising, since the “content” of information coming from the two modalities seems quite different. Nonetheless, the connections between these areas and probably also the frequent correlations in proprioceptive and tactile inputs from same body parts could account for this.

The computational model we presented rests on a number of simplifying assumptions. First, the data set used is preliminary and arbitrary: we used only one coordinated movement type—or ethological category (Affalo and Graziano, 2006)—where the robot was moving the hand in front of its face and following it with gaze. The repertoire of behaviors in newborns needs to be investigated and implemented. Furthermore, newborns’ coordination may in fact be severely limited and it is also possible that the appropriate stimulation type is not coordinated movement, but rather more random and isolated movements such as the muscle twitches (Khazipov et al., 2004). The correlation structure in the proprioceptive data could be induced through the agent’s embodiment. This would probably require a more detailed model of the musculo-skeletal system as well as implementation of reflexes such as the stretch reflex (similar to (Kuniyoshi and Sangawa, 2006)). Second, we did not deal with all the submodalities of proprioception and restricted ourselves to “muscle spindles“—in fact, taking an additional shortcut and directly using joint angles rather than the length of (possibly also biarticular and antagonistically arranged) muscles.

In our model, we experimented with different types of population coding and the Self-Organizing Map algorithm. The most important findings are detailed in what follows. First, as detailed above in Section 2.4.1, Kim et al. (2015) identified different types of proprioceptive neurons in SI, namely position-scaled (single or multi-digit) and posture-selective. It is evident that our model can only learn to represent the posture-selective type. The SOM algorithm is in essence learning to efficiently represent points from the input space, devoting more space in the output layer to more frequent inputs and carrying over topological relationships from the input to the output space. In our case, it basically learns to pick up the most frequent postures (or

“postural synergies“). However, it cannot preserve the position-scaling property—a type of intensity code—increasing neuron firing as a monotonic function of joint position. Finding an algorithm that would preserve this feature while at the same time having the useful properties of SOM (vector quantization, topology preservation) remains the topic of our future work. Second, in our model, the connections from the input space (joint angles, possibly after population coding) to the output layer were all-to-all. That is, every output neuron of the SOM was learning to represent a configuration of all arm and head joints. This is problematic from a mathematical point of view because the topological relationships in the input space—similarity of the different postures in the configurations of all 10 DoF—are clearly of higher dimensionality than 2, while the output layer lattice of the SOM we used was only 2-dimensional and thus the algorithm was faced with the impossible task of reducing the topological manifold from the input space to mere 2 dimensions. Instead, the RFs of neurons in anterior parietal cortex typically span single or multiple joints (Kim et al., 2015), or multiple joints (e.g., 2 or 3) in the posterior parietal cortex (Seelke et al., 2011). Therefore, it seems that the connections between input and output should not be all-to-all but constrained to some regions of the input space only. A modification of the SOM algorithm constraining the maximum receptive field size (MRF-SOM) of the output neurons was presented in (Straka, 2014; Hoffmann et al., 2016). A hierarchy of such maps may give rise to similar gradual expansion of RFs as observed when going upstream in the cortex. Third, we have learned from our study that the SOM algorithm has intrinsic difficulties with input preprocessing using population code. This has to do with the fact that the input (joint angle in this case) is encoded using a population of tuning curves that have different slopes (different in different lines in the linear function case; different even at different points of single curve in sigmoids and Gaussians). After encoding, the SOM weight update occurs in this transformed space; however, this leads to inconsistency in the population if it is to be mapped back to the initial joint space. This is even more problematic if Gaussian curves are used as there is no proper inverse due to their nonmonotonicity. Perhaps, it may not be the necessity for the brain to perform this decoding; yet it seems that the encoding in combination with the SOM adaptation step does introduce undesired transformations to the information that is represented. In the robot, where joint angles are directly available, using them directly leads to superior results. In biological systems, population coding may be inevitable because the information is only available in a distributed fashion, such as from populations of muscle spindles and other receptors. As a consequence, neural learning mechanisms may need to work in different ways than the classical SOM in this case.

Acknowledgment

MH was supported the Marie Curie Intra European Fellowship (iCub Body Schema 625727) within the 7th European Community Framework Programme. We are also indebted to Igor Farkaš for valuable comments and discussions.

Literatúra

- Affalo, T. N. a Graziano, M. S. (2006). Possible origins of the complex topographic organization of motor cortex: reduction of a multidimensional space onto a two-dimensional array. *The Journal of neuroscience*, 26(23):6288–6297.
- Bednárová, N. (2015). Reprezentácia proprioceptívnych vstupov humanoidného robota icub pomocou samoorganizujúcich sa máp]. Diplomová práca, Czech Technical University, Faculty of Electrical Engineering.
- Bergenheim, M., Ribot-Ciscar, E. a Roll, J.-P. (2000). Proprioceptive population coding of two-dimensional limb movements in humans: I. muscle spindle feedback during spatially oriented movements. *Experimental brain research*, 134(3):301–310.
- Bradley, R. M. a Mistretta, C. M. (1975). Fetal sensory receptors. *Physiological Reviews*, 55(3):352–382.
- Butts, D. A. a Goldman, M. S. (2006). Tuning curves, neuronal variability, and sensory coding. *PLoS Biology*, 4(4):e92.
- Cole, J. a Paillard, J. (1995). Living without touch and peripheral information about body position and movement: Studies with deafferented subjects. *The body and the self*, pp. 245–266.
- Drix, D. a Hafner, V. V. (2014). Learning proprioceptive and motor features. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, pp. 374–378. IEEE.
- Feldman, A. (2009). New insights into action–perception coupling. *Experimental Brain Research*, 194:39–58.
- Ghazanfar, A. A., Stambaugh, C. R. a Nicolelis, M. A. (2000). Encoding of tactile stimulus location by somatosensory thalamocortical ensembles. *The Journal of Neuroscience*, 20:3761–3775.
- Hoffmann, M., Straka, Z., Farkas, I., Vavrecka, M. a Metta, G. (2016). Robotic homunculus: Learning of artificial skin representation in a humanoid robot motivated by primary somatosensory cortex. *PLoS ONE [under review]*.

- Huffman, K. J. a Krubitzer, L. (2001). Area 3a: Topographic organization and cortical connections in marmoset monkeys. *Cerebral Cortex*, 11:849–867.
- Jones, K. E., Wessberg, J. a Vallbo, Å. B. (2001). Directional tuning of human forearm muscle afferents during voluntary wrist movements. *The Journal of Physiology*, 536(2):635–647.
- Kandel, E., Schwartz, J. a Jessell, T. (2000). *Principles of Neural Science*, vol. 4. McGraw-Hill.
- Khazipov, R., Sirota, A., Leinekugel, X., Holmes, G. L., Ben-Ari, Y. a Buzsáki, G. (2004). Early motor activity drives spindle bursts in the developing somatosensory cortex. *Nature*, 432(7018):758–761.
- Kim, S. S., Gomez-Ramirez, M., Thakur, P. H. a Hsiao, S. S. (2015). Multimodal interactions between proprioceptive and cutaneous signals in primary somatosensory cortex. *Neuron*, 86(2):555–566.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.
- Krubitzer, L., Huffman, K. J., Disbrow, E. a Recanzone, G. (2004). Organization of area 3a in macaque monkeys: contributions to the cortical phenotype. *Journal of Comparative Neurology*, 471(1):97–111.
- Kuniyoshi, Y. (2014). Learning from examples: Imitation learning and emerging cognition. In *Humanoid Robotics and Neuroscience: Science, Engineering and Society*. CRC Press.
- Kuniyoshi, Y. a Sangawa, S. (2006). Early motor development from partially ordered neural-body dynamics: experiments with a cortico-spinal-musculoskeletal model. *Biol. Cybern.*, 95:589–605.
- Macefield, V. G. (2005). Physiological characteristics of low-threshold mechanoreceptors in joints, muscle and skin in human subjects. *Clinical and Experimental Pharmacology and Physiology*, 32(1-2):135–144.
- Mallot, H. (2013). *Computational Neuroscience: A First Course*. Springer International Publishing.
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C., Rosander, K., Lopes, M., Santos-Victor, J., Bernardino, A. a Montesano, L. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, 23(8-9):1125–1134.
- Milh, M., Kaminska, A., Huon, C., Lapillonne, A., Ben-Ari, Y. a Khazipov, R. (2007). Rapid cortical oscillations and early motor activity in premature human neonate. *Cerebral Cortex*, 17(7):1582–1594.
- Pattacini, U. (2011). *Modular Cartesian Controllers for Humanoid Robots: Design and Implementation on the iCub*. Dizertačná práca, Istituto Italiano di Tecnologia.
- Pellijeff, A., Bonilha, L., Morgan, P. S., McKenzie, K. a Jackson, S. R. (2006). Parietal updating of limb posture: an event-related fmri study. *Neuropsychologia*, 44(13):2685–2690.
- Polani, D. (2002). Measures for the organization of self-organizing maps. Seiffert, U. a Jain, L. (zost.), In *Self-Organizing Neural Networks*, vol. 78 z *Studies in Fuzziness and Soft Computing*, pp. 13–44. Springer.
- Proske, U. a Gandevia, S. C. (2012). The proprioceptive senses: their roles in signaling body shape, body position and movement, and muscle force. *Physiological reviews*, 92(4):1651–1697.
- Saal, H. P. a Bensmaia, S. J. (2014). Touch is a team effort: interplay of submodalities in cutaneous sensibility. *Trends in neurosciences*, 37(12):689–697.
- Sasaki, R., Yamada, Y., Tsukahara, Y. a Kuniyoshi, Y. (2013). Tactile stimuli from amniotic fluid guides the development of somatosensory cortex with hierarchical structure using human fetus simulation. In *Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on*, pp. 1–6.
- Seelke, A. M. H., Padberg, J. J., Disbrow, E., Purnell, S. M., Recanzone, G. a Krubitzer, L. (2011). Topographic maps within brodmann's area 5 of macaque monkeys. *Cerebral Cortex*.
- Shadmehr, R. a Wise, S. P. (2005). *The computational neurobiology of reaching and pointing: a foundation for motor learning*. MIT press.
- Sherrington, C. (1906). On the proprioceptive system, especially in its reflex aspects. *Brain*, 29:467–482.
- Straka, Z. (2014). Tvorba senzorických map z taktilních vstupů robota iCub pomocí metod strojového učení. Diplomová práca, Czech Technical University, Faculty of Electrical Engineering.
- Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L. a Nori, F. (2008). An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, pp. 57–61. ACM.
- Vesanto, J., Himberg, J., Alhoniemi, E. a Parhankangas, J. (1999). Self-organizing map in Matlab: the SOM toolbox. In *Proceedings of the Matlab DSP Conference*, vol. 99, pp. 16–17.

Yamada, Y., Fujii, K. a Kuniyoshi, Y. (2013). Impacts of environment, nervous system and movements of preterms on body map development: Fetus simulation with spiking neural network. In *Development and Learning and Epigenetic Robotics (ICDL), 2013 IEEE Third Joint International Conference on*, pp. 1–7.

Žabkar, J. a Leonardis, A. (2016). Motor memory: Representation, learning and consolidation. *Biologically Inspired Cognitive Architectures*, 16:64–74.

Zipser, D. a Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331(6158):679–684.

Funkčné zmeny mozgu počas spracovávania sociálnych informácií u schizofrénie: multimodálna meta-analýza neurozobrazovacích štúdií.

Martin Jáni, Tomáš Kašpárek

CEITEC - Stredoevropský technologický inštitút, Masarykova Univerzita
Kamenice 5, 625 00, Brno, Česká republika
Psychiatrická klinika Lékařské fakulty Masarykovy univerzity a Fakultní nemocnice Brno
Lékařská fakulta, Kamenice 5, 625 00, Brno, Česká republika
e-mail: martinjani@mail.muni.cz, tomas.kasperek@centrum.cz

Abstrakt

Neurozobrazovacie, tak aj behaviorálne štúdie uvádzajú nejednoznačné výsledky ohľadom spracovávania emočných a kognitívnych mentálnych stavov u schizofrénie. Neurozobrazovacie štúdie skúmajúce emočnú rekogníciu a teóriu mysle sme analyzovali pomocou metódy Seed-based d mapping. Pacienti vykazovali abnormálnu aktiváciu v oboch modalitách, čo naznačuje komplexný deficit sociálnej kognície na neurálnej úrovni.

1 Úvod

Sociálna kognícia predstavuje kognitívne procesy, ako napr. pracovná pamäť alebo vnímanie, ktoré nám pomáhajú porozumieť a predvídať správanie druhých. Správna interpretácia zámerov a porozumenie emóciám sú totiž nevyhnutné pre úspešné fungovanie v spoločnosti. Schizofrénia je charakteristická ťažkosťami v interpersonálnom fungovaní a bolo zistené, že tieto ťažkosti u pacientov súvisia viac s úrovňou sociálnej kognície, než neurokognície (Fett et al., 2011). Narušenie sociálnej kognície u schizofrénie dokladajú početné výskumy: zisťuje sa, že sociálna kognícia je do istej miery narušená u príbuzných prvého stupňa, rodičov a súrodencov, u pacientov so schizofréniou môžeme zachytiť deficit už v premorbídnom štádiu, a pretrváva aj po odznení psychotických príznakov (Mehta et al., 2013; Modinos et al., 2010). Všetky tieto zistenia naznačujú neurobiologický pôvod narušenia sociálnej kognície, podľa Greena a spol. (2015) tak z nej robia vhodného kandidáta na endofenotyp pre schizofréniu (endofenotypy su kvantitatívne markery rysovej povahy a vypovedajú o genetickom základe ochorenia). Sociálna kognícia zahŕňa viacero procesov, medzi najviac preskúmané domény sociálnej kognície u schizofrénie patrí rozpoznávanie emócií v tvári (emočná rekognícia) a schopnosť rozpoznať a

porozumieť predstavám, zámerom a myšlienkam druhých (Teória mysle).

U pacientov so schizofréniou boli zistené ťažkosti v rozpoznávaní emócií v tvárach druhých (Barkl et al., 2014), sú tiež viac senzitivní k špecifickým emóciám ako strach alebo hnev (Mandal et al., 1998). Rovnako zdraví jedinci majú väčšie ťažkosti s rozpoznávaním emócií z výrazu tváre u pacientov so schizofréniou. To môže súvisieť s celkovým komplexným narušením spracovávania emócií u schizofrénie, vyznačujúcim sa oploštelou emotivitou, zvýšenou alexithymiou (identifikovania a popisovania emócií), inklinovaním k používaniu neadaptívnych emočne regulatívnych mechanizmov – potlačením (Kimhy et al., 2012; Meer et al., 2009) a nižším emočným uvedomením (Baslet et al., 2009). Podobne, ťažkosti s teóriou mysle boli u pacientov so schizofréniou opakovane doložené (Bora et al., 2009; Sprong et al., 2007). I keď ťažkosti so spracovaním sociálnych informácií sa zistili v oboch doménach, vzťah týchto ťažkostí a teda samotnej povahy ochorenia je stále nejasné. U niektorých experimentálnych štúdií môžeme vidieť rozlíšenie afektívnej a kognitívnej Teórie mysle – tj. rozlíšením medzi porozumením afektívnym stavom a porozumením kognitívnym stavom druhých. Prvý menovaný proces je nazývaný aj afektívna teória mysle (ako komponent teórie mysle), alebo tiež kognitívna empatia (ako komponent empatie), čo spôsobuje značný terminologický zmätok. Napovedá nám však o dôležitosti prepojenia afektívnej a kognitívnej zložky u sociálnej kognície, i keď štúdie porovnávajúce tieto zložky nám prinášajú rozporuplné výsledky, snáď aj kvôli konceptuálnym odlišnostiam. Napr. štúdie, kde sa zistila narušená kognitívna zložka ale neporušená afektívna, používajú komiksové úlohy (Benedetti et al., 2009; Lee et al., 2010), kde treba odvodiť nasledujúce správanie z kontextu, testovaný proces tak môže zahŕňať vedomosti o sociálnych normách a konvenciách. Na druhej strane u štúdií, ktoré našli deficit afektívnej zložky sa užíva tzv. Yoni task (Shamay-Tsoory et al., 2007), založená na odhadovaní mentálnych stavov podľa smeru pohľadu, a teda umožňuje zjednodušujúce mechanické riešenie.

Najnovší výskum používajúci filmové stimuly na komplexné posúdenie sociálnej kognície našiel narušenie oboch, tak kognitívnej, ako aj afektívnej zložky (Montag et al., 2011).

Lepšie porozumenie vzťahu medzi narušením kognitívnej a afektívnej zložky sociálnej kognície u schizofrénie by nám mohli priniesť neurovedné štúdie. Určité oblasti v mozgu tvoria sieť, ktorá sa zapája počas spracovávania sociálnych informácií, táto sieť zahŕňa mediálny prefrontálny kortex, predné cingulum, temporo-parietálny spoj, precuneus a sulcus temporalis superior (Schurz et al., 2014). Ďalšie, hlavne podkorové oblasti sa zapájajú pri spracovávaní emócií u druhých, a pokiaľ ide spracovávanie emočných výrazov v tvári, je zapojený aj gyrus fusiformis, oblasť tradične spájaná s rozpoznávaním tvári. U schizofrénie sa však zistilo, že počas tvárovej emočnej rekognície sa tie posledne menované oblasti aktivujú z menšej miere, na druhej strane insula, temporálne a parietálne oblasti sa aktivujú viac v porovnaní so zdravými kontrolami. Green a spol. (2015) to môže značiť, že pacienti majú narušené mozgové obvody zodpovedné za spracovávanie emócií v tvárach a preto to kompenzujú inými kognitívnymi procesmi, pričom hyperaktivované oblasti naznačujú na teóriu mysle. Aby sme to však mohli overiť, je potrebné vedieť, ako spracovávajú mentálne stavy druhých pacienti so schizofréniou na neurálnej úrovni. Doterajšie výsledky sa zhodujú v tom, že pacienti vykazujú abnormálnu aktiváciu, avšak nie sú konzistentné, čo sa týka siete aktivácie. Preto sme previedli meta-analýzu neurozobrazovacích štúdií teórie mysle a následne emočnej rekognície u schizofrénie. Výskum teda poskytuje ucelený kvantitatívny prehľad neurálnych substrátov sociálnej kognície u tohto ochorenia.

2 Metóda

29 neurozobrazovacích štúdií (fMRI a PET), ktoré zahŕňali explicitnú úlohu testujúcu teóriu mysle (n = 16) alebo emočnú rekogníciu (n = 13), bolo zahrnutých do do analýzy. Na analýzu sme použili SDM softvér (Seed-based d mapping, predtým Signed Differential Mapping)(Radua et al., 2012; www.sdmproject.com). SDM je metóda, ktorá stavia na pozitívnych vlastnostiach starších meta-analytických metódach ako Activation Likelihood Estimation (ALE)(Eickhoff et al., 2009) alebo Multilevel Kernel Density Analysis (MKDA)(Wager et al., 2007), ale pridáva nové prvky. Vytvorí parametrické mapy síl efektov a mapu variancie, odvodenú od sily efektu a veľkosti vzorku štúdie.

3 Výsledky a diskusia

Ako bolo zistené už u predchádzajúcich štúdií, pacienti so schizofréniou počas emočnej rekognície zapájajú

v zníženej miere oblasti zodpovedné za spracovávanie tvárových informácií, a to ventrolaterálny prefrontálny
Emočná rekognícia



Teória mysle



Multimodálna meta-analýza



■ HC < SZ
■ HC > SZ
■ HC > SZ počas ER a SZ > HC počas ToM

Ob. 1: Výsledky meta-analýzy. SZ = schizofrénia, HC = zdravé kontroly, ER = emočná rekognícia, ToM = teória mysle.

kortex, cingulum, insulu a subkortikálne oblasti zahŕňajúce amygdalu, thalamus, nucleus caudatus, nucleus lentiformis a putamen. Na druhej strane pacienti so schizofréniou vykazovali zvýšenú aktiváciu v parietálnom kortexe (inferiorný parietálny lalok) v porovnaní so zdravými kontrolami.

Pacienti vykazovali abnormálnu aktiváciu počas úloh ohľadom teórie mysle v oblastiach, ktoré sa tradične spájajú s týmito procesmi, konkrétne v pravom superiornom temporálnom gyre, insule, kuneu, mediálnom prefrontálnom kortexe a bilaterálne v parietálnom kortexe. U všetkých oblastí vykazovali pacienti zníženú aktiváciu v porovnaní so zdravými kontrolami, s výnimkou posledne menovaných oblastí parietálnych lalokov.

Pomocou multimodálnej meta-analýzy sme zisťovali prekryv abnormálnej aktivácie pacientov so schizofréniou oproti zdravým kontrolám emočnej rekognície a teórie mysle. Znížená aktivácia spoločná pre obe modalítity sa opakovala u pravej insuly, superiornom temporálnom sulku a anteriornom cingule zasahujúcom do mediálneho prefrontálneho kortexu. Zvýšenú aktiváciu sme pozorovali u parietálnych oblastí, zahŕňajúcich supramarginálny gyrus. Pravý inferiorný frontálny gyrus vykazoval zvýšenú aktiváciu počas teórie mysle a zároveň zníženú aktiváciu počas emočnej rekognície u schizofrénie oproti zdravým kontrolám.

Tieto výsledky naznačujú komplexný deficit spracovávania sociálnych informácií naprieč modalitami. Aktivácia pravého superiorného temporálneho laloku je spájaná s detekciou biologického pohybu (Han et al., 2013), zatiaľ čo s mediálnym prefrontálnym kortexom sa spája spracovávanie trvácnejších sociálnych informácií, ako

sociálnych noriem alebo usudzovanie ohľadom charakterových čŕt (Van Overwalle, 2009). Anteriórne cingulum spolu s insulou tvorí sieť, ktorá slúži na spracovávanie sensorických a emočných informácií. Táto sieť je dôležitá pre interopeciu a jej narušenie môže mať následky súvisiace s reprezentáciou samého seba (White et al., 2010). Je možné, že práve s tým súvisí zvýšená aktivácia parietálnych oblastí, zvlášť supramarginálneho gyru, ktorý hraje kľúčovú rolu v potlačení vlastnej perspektívy, a umožňuje tak zaujať perspektívu druhého (Silani et al., 2013). To by mohol byť spôsob, ktorým pacienti so schizofréniou kompenzujú deficit inými kognitívnymi stratégiami, napr. prehnaným sebamonitorovaním. Tiež je však možné, že prepínať medzi perspektívou seba a druhých je pre pacientov so schizofréniou viac náročné, čo by sa v tomto prípade prejavilo zvýšenou neurálnou aktivitou.

Meta-regresná analýza ukázala, že aktivácia pravého inferiórneho frontálneho gyru a kuneu súvisí s dĺžkou ochorenia a aktivácia superiórneho temporálneho gyru a precentrálneho gyru súvisí s vyššími dávkami antipsychotík. Pozitívne symptómy súviseli aktiváciou superiórneho temporálneho gyru a kuneu, zatiaľ čo závažnejšie negatívne symptómy boli asociované s nižšou aktiváciou pravého supramarginálneho gyru.

4 Zhrnutie

Pomocou SDM metódy sme zistili abnormálnu aktiváciu u schizofrénie v oblastiach zodpovedných za spracovávanie emočných informácií v tvárach. Tak isto sme zistili abnormálnu aktiváciu v sieti mozgových oblastí spojených s usudzovaním na mentálne stavy druhých. Multimodálna meta-analýza ukázala zníženú aktiváciu v pravej insule, temporálnom kortexe, anteriórnom cingule a mediálnom prefrontálnom kortexe v oboch modalitách u schizofrénie, zároveň však zvýšenú aktiváciu v parietálnom kortexe. Výsledky naznačujú komplexný deficit sociálnej kognície na neurálnej úrovni. Meta-regresná analýza ukázala, že dĺžka ochorenia, medikácia a symptomatológia môžu mať vplyv na narušenú sieť sociálnej kognície u schizofrénie.

PodĎakovanie

Tento príspevok vznikol za podpory MŠMT v rámci projektu Implementace behaviorálních paradigmat pro zobrazování a volbu neurostimulačních protokolů v psychiatrii (MUNI/A/1240/2014).

Literatúra

Barkl, S. J., Lah, S., Harris, A. W. F., & Williams, L. M. (2014). Facial emotion identification in early-

onset and first-episode psychosis: a systematic review with meta-analysis. *Schizophrenia Research*, 159(1), 62–9.

Baslet, G., Termini, L., & Herbener, E. (2009). Deficits in emotional awareness in schizophrenia and their relationship with other measures of functioning. *The Journal of Nervous and Mental Disease*, 197(9), 655–660.

Benedetti, F., Bernasconi, A., Bosia, M., Cavallaro, R., Dallspezia, S., Falini, A., ... Smeraldi, E. (2009). Functional and structural brain correlates of theory of mind and empathy deficits in schizophrenia. *Schizophrenia Research*, 114(1-3), 154–60.

Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia Research*, 109(1-3), 1–9.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, 30(9), 2907–26.

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-G., Penn, D. L., van Os, J., & Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neuroscience and Biobehavioral Reviews*, 35(3), 573–588.

Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews. Neuroscience*, 16(10), 620–631.

Han, Z., Bi, Y., Chen, J., Chen, Q., He, Y., & Caramazza, A. (2013). Distinct regions of right temporal cortex are associated with biological and human-agent motion: functional magnetic resonance imaging and neuropsychological evidence. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(39), 15442–53.

Kimhy, D., Vakhrusheva, J., Jobson-ahmed, L., Tarrier, N., Malaspina, D., & Gross, J. J. (2012). Emotion awareness and regulation in individuals with schizophrenia: Implications for social functioning. *Psychiatry Research*, 200(2-3), 193–201.

Lee, S. J., Kang, D. H., Kim, C.-W. W., Gu, B. M., Park, J.-Y. Y., Choi, C.-H. H., ... Kwon, J. S. (2010). Multi-level comparison of empathy in schizophrenia: an fMRI study of a cartoon task. *Psychiatry Research*, 181(2), 121–9.

- Mandal, M. K., Pandey, R., & Prasad, a B. (1998). Facial expressions of emotions and schizophrenia: a review. *Schizophrenia Bulletin*, 24(3), 399–412.
- Meer, L. Van Der, Van, M., & Aleman, A. (2009). Emotion regulation strategies in patients with schizophrenia. *Psychiatry Research*, 170(2-3), 108–113.
- Mehta, U. M., Thirthalli, J., Naveen Kumar, C., Keshav Kumar, J., Keshavan, M. S., & Gangadhar, B. N. (2013). Schizophrenia patients experience substantial social cognition deficits across multiple domains in remission. *Asian Journal of Psychiatry*, 6(4), 324–9.
- Modinos, G., Renken, R., Shamay-Tsoory, S. G., Ormel, J., & Aleman, A. (2010). Neurobiological correlates of theory of mind in psychosis proneness. *Neuropsychologia*, 48(13), 3715–24.
- Montag, C., Dziobek, I., Richter, I. S., Neuhaus, K., Lehmann, A., Sylla, R., ... Gallinat, J. (2011). Different aspects of theory of mind in paranoid schizophrenia: Evidence from a video-based assessment. *Psychiatry Research*, 186(2-3), 203–209.
- Shamay-Tsoory, S. G., Shur, S., Barcai-Goodman, L., Medlovich, S., Harari, H., & Levkovitz, Y. (2007). Dissociation of cognitive from affective components of theory of mind in schizophrenia. *Psychiatry Research*, 149(1-3), 11–23.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 42(0), 9–34.
- Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(39), 15466–76.
- Sprong, M., Schothorst, P., Vos, E., Hox, J., & van Engeland, H. (2007). Theory of mind in schizophrenia: meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, 191, 5–13.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858.
- Wager, T. D., Lindquist, M., & Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience*, 2(2), 150–8.
- White, T. P., Joseph, V., Francis, S. T., & Liddle, P. F. (2010). Aberrant salience network (bilateral insula and anterior cingulate cortex) connectivity during information processing in schizophrenia. *Schizophrenia Research*, 123(2-3), 105–15.

Vyššie emócie: neurovedecký a psychologický pohľad

Dalibor Jurášek
Peter Marman

Univerzita Komenského v Bratislave
Filozofická fakulta
Gondova 2
814 99, Bratislava
marman@fphil.uniba.sk, jurasek.dalibor@gmail.com

Abstrakt

S konceptom vyšších emócií sa v našej psychologickej literatúre stretávame zriedka a v zahraničnej takmer vôbec. Existujúce kategorizácie a teórie emocionalitu vnímajú ako v zásade jednorovňový proces. Poznatky zo štúdií mozgu však ukazujú, že emocionalita je v mozgu sprostredkovaná viacerými oblasťami rôznej fylogenetической zložitosti. V našom článku stručne popisujeme tieto neurovedecké štúdie a na základe poznatkov z viacerých disciplín načrtáme psychologický koncept vyšších emócií, ktorý by sa mohol stať súčasťou budúcej integratívnej teórie. Uvádzame aj odporúčania pre výskum a prax.

1 Úvod

Členenie človeka na „nižšiu“ a „vyššiu“ zložku je odvekou predstavou. Od počiatkov filozofie sa diskutuje o „ľudskej prirodzenosti“ a o tom, či v človeku je čosi „vyššie“, kvalitatívne odlišné od biologického, „zvieracieho“. Aj dnes, v sekulárnej spoločnosti, rezonuje predstava, že človek by sa mal snažiť na sebe pracovať, rozvíjať svoju ľudskosť a neradiť sa iba bezprostrednými pudmi a potrebami.

Čo nám na túto tému hovorí psychológia? V oblasti myslenia a konania rozlišuje zhruba dva protipóly: myslenie viac intuitívne, nevedomé, rýchle a myslenie viac racionálne, vedomé, pomalšie (Kahneman, 2013); pri správaní automatizované a kontrolované úkony a proces automatizácie (Sternberg, 2002). Objavujú sa dokonca aj výskumy vedomej vôle a jej pôsobenia späť (Jurášek, Marman, 2013).

Keď sa však pozrieme na zložku emócií, členenie na vyššie a nižšie prakticky chýba. Kapitoly o vyšších emóciách sú ešte u niektorých našich autorov (Kubáni, 2010; Nakonečný, 2012), v západných publikáciách sme ich ale nenašli vôbec, či už v prehľadových učebniciach (Atkinson et al., 2003; Kassin, 2007; Hill, 2004) alebo teóriách emócií a emočných klasifikáciách (Plutchik,

2002; Robinson, 2008; Baron-Cohen, 2004; Izard, 1991; Goleman, 1997).

Na druhej strane neuroveda sa venuje výskumu aj tzv. seba-reflektujúcich emócií. Tieto emócie aktivujú vyššie časti kortexu, vyžadujú pokročilé seba-vedenie a majú vzťah k etickým alebo iným normám. Existencia seba-reflektujúcich emócií, kvalitatívne odlišných od základných emócií, sa všeobecne uznáva. Na druhej strane tieto poznatky nie sú zapracované do psychologickej teórie – neexistuje žiadny model, ktorý by ucelene popisoval ich fungovanie. V tomto článku chceme stručne predstaviť výsledky neurovedeckého skúmania seba-reflektujúcich emócií, načrtnúť náš psychologický koncept a uviesť odporúčania pre výskum a prax.

2 Neurovedecké štúdie

2.1 Základné emócie

Časť neurovedeckých výskumov skúma koreláty emočného prežívania. Stav poznania v tejto oblasti – viac ako 15 rokov bádania – už umožnil vypracovanie viacerých metaanalýz (Phan et al., 2002; Vytal, Hamann, 2010; Lindquist et al., 2012). Veľká väčšina štúdií, ako i všetky metaanalýzy, sa zaoberajú základnými emóciami ako hnev, strach alebo radosť. Z týchto štúdií emerguje klasický konsenzus, že emócie súvisia hlavne s limbickým systémom (Stuchlíková, 2002). Časti kortexu plnia rolu pri rozhodovaní (Bechara, Damasio, Damasio, 2000) a regulácii emócií (Beauregard, Lévesque, Bourgouin, 2001), oveľa menej pri samotnom emočnom prežívaní.

2.2 Seba-reflektujúce emócie

Podstatne menšie množstvo výskumov je venované iným ako základným emóciám. Ide o emócie, ktoré vyžadujú vysoko vyvinutý koncept Self, často majú sociálny charakter a súvis s etickými, resp. inými normatívnymi systémami. Nazývajú sa rôzne:

- seba-reflektujúce emócie (Roth et al., 2014),
- seba-vedomujúce emócie (Leary, 2004; Tracy, Robins, 2004; Gilead et al., 2015),
- morálne emócie (Fourie et al., 2014; Fox et al., 2015),
- sociálne emócie (Basile et al., 2011),
- emócie druhého rádu (Hammond, 2006).

Ďalej budeme jednotne používať výraz seba-reflektujúce emócie, resp. vyššie emócie pri náčrte nášho konceptu, pričom pod nimi rozumieme aj všetky vyššie spomenuté.

2.3 Mozgové koreláty seba-reflektujúcich emócií

S ktorými oblasťami mozgu sú tieto emócie späté? Fourie et al. (2014) skúmali mozgové koreláty pocitu viny a zistili zvýšené aktivácie v cingulárnej kôre, prednej inzule, dorzomediálnom prefrontálnom kortexe a prekuneu. Kédia et al. (2008) skúmali mozgové koreláty štyroch seba-reflektujúcich emócií a zistili zvýšené aktivácie v dorzomediálnom prefrontálnom kortexe, prekuneu a spánkovo-parietálnom spoji. Basile et al. (2011) našli pri pocitoch viny zvýšenú aktiváciu v cingulárnom kortexe a mediálnom frontálnom závite.

Fox et al. (2015) našli pri pocite vďačnosti zvýšené aktivácie v prednej cingulárnej kôre a mediálnom prefrontálnom kortexe. Gilead et al. (2015) porovnávali emócie viny a hanby so základnými emóciami a našli zvýšené aktivácie v prípade viny a hanby v mediálnom prefrontálnom kortexe a v laterálnom dorzálnom prefrontálnom kortexe. Roth et al. (2014) skúmali mozgové aktivácie pri pýche a vine, ktoré vyšli veľmi podobne a zahŕňali klasické emočné oblasti ako amygdala, inzula a ventrálne striátum, a tiež dorzomediálny prefrontálny kortex.

Doterajšie výsledky teda ukazujú, že aj keď základné emócie niekedy aktivujú časti kortexu a seba-reflektujúce emócie niekedy aktivujú aj podkôrové štruktúry, pri seba-reflektujúcich emóciách je jasne častejšia aktivácia kortexu, predovšetkým prefrontálneho kortexu. Aj keď ešte neexistuje metaanalýza mozgových korelátov seba-reflektujúcich emócií, tento záver bol potvrdený každou zo spomínaných štúdií.

Psychologické procesy úzko súvisia s mozgovými procesmi. Preto skutočnosť, že seba-reflektujúce emócie aktivujú iné oblasti mozgu, by sa mala odraziť aj v psychológii. Tieto poznatky však zatiaľ nie sú integrované do komplexnejšieho kauzálneho modelu. Ako sme spomínali, všetky významné teórie a klasifikácie emócií rátajú s emocionalitou ako jednoúrovňovým procesom (Plutchik, 2002; Izard, 1991; Robinson, 1998; Baron-Cohen, 2004). Pokúsime sa preto načrtnúť náš psychologický koncept, ktorý prepojíme s neurovedeckým poznaním.

3 Vyššie emócie: náčrt psychologického modelu

3.1 Ciele teórie emócií

Pod teóriou a klasifikáciou emócií nevnímame iba deskriptívny model, ale predovšetkým stanovenie vhodných základných konceptov a kauzálnych prepojení medzi nimi. Nestačí popísať, aké emócie existujú a kedy vznikajú, ale je potrebné v prvom rade objasniť súvislosti: *prečo* vznikajú a ako, ak vôbec, ich možno meniť.

Komplexná teória emócií by mala preto byť schopná:

- **Charakterizovať emócie:** vysvetliť, čo to emócie sú, pričom táto charakteristika by sa mala vzťahovať na všetky emócie (nielen základné).
- **Kategorizovať emócie:** sú niektoré emócie spätéjšie v tom zmysle, že je za nimi rovnaká alebo podobná psychická dynamika?
- **Popísať vývin emócií:** človek sa rodí do stavu psychiky, v ktorom má v podstate len nevedomie a základné emócie. V priebehu života môže dosiahnuť vysoké úrovne myslenia a komplexných seba-vedomujúcich emócií. Ako dochádza k tejto zmene? Menia sa niektoré základné emócie na vyššie emócie?
- **Dať odporúčania do praxe:** je možné zmeniť emočné vzorce vedomie? Možno určitú základnú emóciu cielene pretransformovať na vyššiu emóciu? Ak áno, ktorú na ktorú a ako to docieľiť? Tieto otázky sú základnými otázkami pre terapiu a poradenskú psychológiu.

V tomto článku nechceme popísať celú našu teóriu emócií (na ktorej pracujeme), ale poukázať na dôležitý aspekt: existenciu nižších a vyšších emócií. Vychádzame z existujúceho poznania a dopĺňame ho vlastnými psychologickými pozorovaniami.

3.2 Dimenzie vyšších emócií

Poznatky z neurovedy, evolučnej biológie, fyziológie, štúdia kultúry a viacerých oblastí psychológie nasvedčujú, že existujú dva typy, resp. presnejšie dva protipóly emócií – základné a vyššie.

Neurovedecká dimenzia: vyššie emócie vo väčšej miere ako základné aktivujú kôrové časti mozgu, predovšetkým prefrontálneho kortexu. Toto pozorovanie je konzistentné s už uvedenými výskumami neurológie vyšších emócií (Kédia et al., 2008; Fourie et al., 2014; Basile et al., 2011; Fox et al., 2015; Roth et al., 2014).

Biologická dimenzia: základné emócie sa v evolúcii objavili podstatne skôr, najpravdepodobnejšie už

u prvých zvierat. Aj dnes možno v etológii pozorovať mnohé emočné prejavy zvierat zodpovedajúce základným emóciám (Reimerta, 2013; Rutherford, 2012). U vyšších primátov nájdeme aj náznaky vyšších emócií, napr. v empatii (Warneken et al., 2007; Campbell, de Waal, 2014) a plné rozvinutie vyššie emócie dosahujú u človeka. Iné zvieratá, aj keď sa u nich môže objavovať altruistické správanie, nemajú dostatočný stupeň sebauvedomenia, aby sa dalo hovoriť o vyšších emóciách v pravom slova zmysle. Objavenie sa základných emócií, ako aj postupný rozvoj vyšších emócií prebiehal, pochopiteľne, plynule.

Fyziologická dimenzia: základné emócie sú úzko späté s fyziológiou tela a prejavujú sa v telesných procesoch (tlkot srdca, tlak, potenie, dýchanie, teplota kože) a tvárových výrazoch (Keltner et al., 2003; Ekman, 2003). Vyššie emócie, ako napríklad pokora, úcta, vďačnosť nemajú také zjavné telesné prejavy a nepodarilo sa identifikovať ani ich spoločné tvárové výrazy (Tracy, Robins, 2004).

Kulturologická dimenzia: základné emócie sa vyskytujú medzikultúrne a vedia ich rozpoznať aj jedinci z rozličných kultúr, napr. v tvárových výrazoch (Ekman, 2003). Vyznačujú sa aj charakteristickou gestikou, čo môžeme dobre vidieť v divadle. Naproti tomu, spoznať v tvári a gestách vyššie emócie ako spravodlivosť, kajúcnosť, nádej alebo dôvera, či dokonca ich zahrať, je výrazne náročnejšie.

Psychologická dimenzia: z hľadiska vývinovej psychológie sa všetky základné emócie dajú u dieťaťa pozorovať už v predškolskom veku. Vyššie emócie sa objavujú postupne a úzko súvisia s vývinom sebauvedomenia. Na mnohé je ale potrebné pochopenie komplexných myšlienkových konštruktov, takže sa objavujú v plnej miere až v dospelosti.

Z hľadiska vzťahu k ľudskému Ja je človek pri základných emóciách vnútorne pasívnejší, majú tendenciu aktivovať sa a doznieť nevedome. Na vyšších emóciách je však potrebné cielene pracovať a vedome ich vyvíjať, hlavne čo sa týka vyšších stupňov ich rozvinutia (do istej miery sa vyššie emócie dostávajú do človeka aj socializáciou a na základe vrodenných predpokladov; až pre ich silnejšie rozvinutie a plne vedomé uchopenie je potrebné cielene vedome pracovať).

Dôležitým aspektom vyšších emócií je sýtenie kongitívnym elementom. Základné emócie majú malú spätosť s myslením. Vyskytujú sa aj u zvierat, úplne malých detí, rovnako u ľudí so subnormnými kognitívnymi schopnosťami a dajú sa u nich jasne rozoznať. U vyšších emócií je zásadný a nevyhnutný istý myšlienkový koncept v zmysle appraisal teórií (Ellsworth, Scherer, 2003), ktorý môže byť vysoko rozvinutý a sofistifikovaný (aj keď nemusí byť vedomý).

Hlavné dimenzie rozdielov medzi základnými a vyššími emóciami zhŕňame v tabuľke.

| Základné emócie | Vyššie emócie |
|---|--|
| Fylogeneticky staršie | Fylogeneticky mladšie |
| Prítomné aj u zvierat | Špecificky ľudské, u zvierat len naznačené |
| Silná spätosť s fyziológiou, látkami v tele | Menšia spätosť s fyziológiou, látkami v tele |
| Jasne charakteristické tvárové výrazy a gestika | Menej jasne charakteristické tvárové výrazy a gestika |
| V mozgu menej späté s neokortexom | V mozgu viac späté s neokortexom |
| V individuálnom vývine sa objavujú už v detstve | V individuálnom vývine sa objavujú neskôr, rozvíjajú sa celý život |
| Menej racionálne a vedomé | Viac racionálne a vedomé |
| Kognitívne sýtenie je slabé | Kognitívne sýtenie je zásadné, obsahujú myšlienkovú schému |
| Lokus kontroly: človek je vnútorne pasívnejší | Lokus kontroly: človek je vnútorne aktívnejší |
| Kultúrne univerzálne | Kultúrne špecifickejšie |
| Predstavujú biologický základ | Predstavujú nadstavbu nad biologickým základom |
| Príklady: strach, hnev, hnus, radosť... | Príklady: pokora, dojatie, pocit krásna, úcta, súciti... |

Tab. 1: Rozdiely medzi základnými a vyššími emóciami.

Upresnili sme charakteristiku vyšších emócií a zahrnuli sme ich do kontextu viacerých disciplín. Pre teóriu emócií sú ale dôležité aj súvislosti a kauzálne princípy, ktoré im podliehajú.

3.3 Súvislosti vyšších emócií

3.3.1 Počet vyšších emócií

Neurovedecké výskumy vyšších emócií sa obvykle orientujú na vinu, hanbu a pýchu. Zopár štúdií je zameraných na obdiv alebo vďačnosť. Vyšších emócií je však oveľa viac. Už len analýza slovnej zásoby (Marman, Jurášek, v príprave) ukazuje desiatky vyšších emócií ako dojatie, dôvera, hrdosť, kajúcnosť, krásno, nádej, odvaha, pieta, súciti, vďačnosť a iné. Aby sme pochopili ľudskú emocionalitu, je potrebné zaoberať sa aj týmito emóciami a zahrnúť ich do teórie emocionality.

3.3.2 Etický rozmer

Pri vyšších emóciách sa objavuje dôležitý rozmer, a tým je etika. Slovo „vyššie emócie“ sa dokonca niekedy rezervuje len pre tieto etické emócie. Kľúčovým kritériom je tu postoj ľudského Ja. Pri etických emóciách je ľudské Ja vymedzené pozitívne, sympaticky: otvorené

a prijímajúce k sebe aj druhým a je pripravené svoju energiu investovať do konštruktívnej práce, prijať svoju rolu v spoločnosti a prispieť do nej (pokora, vďačnosť, súcit, odvaha). Pri neetických emóciách je ľudské Ja vymedzené antipaticky: buď proti sebe, čo mu znemožňuje otvoriť sa druhým a paralyzuje ho, takže je neschopné prijať svoju spoločenskú úlohu (vina, hanba); alebo antipatiu obracia proti druhým a stavia sa proti nim (závisť, pýcha, pohrdanie).

Pochopiteľne, reálne zažívané emócie sa vyskytujú v rôznych prechodoch a kombináciách. Treba si byť vedomý aj obmedzenia slovných kategórií, napr. pod hrdosťou, pýchou alebo láskou sa môžu myslieť rôzne veci; navyše jeden slovný koncept môže zahŕňať viacero emočných procesov. Pri konštruovaní teórie je potrebné si všimnúť konkrétnu emóciu, nie limitovať sa slovným vyjadrením.

3.3.3 Metamorfóza emócií

Je dôležité popísať premeny emócií. Človek sa rodí v stave, v ktorom má prakticky len základné emócie a len postupne si formuje vlastné Ja, osobnosť a získava vôbec kapacitu vyššie emócie zažívať. Pridávajú sa vyššie emócie ako čosi navyše, alebo vznikajú kvalitatívnou transformáciou nižších? K zmenám emočných vzorcov navyše dochádza nielen vonkajšími podnetmi a zrením, ale aj vnútornými úkonmi (napr. meditácia, psychoterapia, modlitba). Dajú sa nižšie emócie pretaviť na vyššie? Ak áno, ako? A ktoré sa premienia na ktoré?

3.3.4 Interdisciplinárne súvislosti

Popísanie vyššie uvedených súvislostí je potrebné na porozumenie ľudským emóciám. Zároveň je dôležité sledovať kauzálne linky interdisciplinárne: v biológii (evolúcia emócií, porovnanie človeka a zvierat), medicíne (somatizácia, vzťah emočných vzorcov k imunitě a chronickým chorobám), teoretickej i aplikovanej psychológii (terapia, výcvik), kultúre (umenie, tradície, zvyky) a v neposlednom rade vo filozofii (kam emócie smerujú, majú nejaký hlbší zmysel?). Ucelená teória by mala dávať prepojený obraz a odporúčania do praxe.

4 Záver

V súčasných najvplyvnejších psychologických teóriách nenachádzame rozlíšenie emócií na základné a vyššie. Konvergujúce poznanie z neurovedy, evolučnej biológie, fyziológie, štúdia kultúry a psychológie však ukazuje, že existujú dva protipóly emócií: emócie základné (alebo nižšie): viac späté s limbickým systémom, v evolúcii sa objavujú skôr, majú jasnú odozvu vo fyziológii, látkach v tele, tvárových výrazoch a gestike, sú kultúrne univerzálne, v individuálnom

vývine sa objavujú skôr a sú viac nevedomé a iracionálne; a emócie vyššie: viac späté s neokortexom, plne sa objavujú až u človeka, majú menšiu odozvu vo fyziológii, látkach v tele, tvárových výrazoch a gestike, sú kultúrne špecifickejšie, v individuálnom vývine sa objavujú neskôr a sú viac vedomé a racionálne.

Okrem popísania konceptu vyšších emócií je dôležité uvedomiť si, že ich je oveľa viac ako emócií základných (rádovo desiatky); že sa v nich objavuje etický rozmer daný postojom Ja (etický, sympatický, prosociálny vs. neetický, antipatický, antisociálny); a že môže dochádzať k metamorfóze emócií: vzorce základných emócií, resp. vyšších neetických emócií sa dajú pretaviť na vyššie etické emócie.

Súvislosti základných a vyšších emócií sa dajú sledovať multidisciplinárne: v psychológii, biológii, medicíne, kultúre a filozofii. Porozumenie týmto súvislostiam je predpokladom emočnej teórie, ktorá je komplexná a umožňuje dať odporúčania do praxe – bežného života a terapie. Ďalší výskum týchto otázok by mal tvoriť základ integratívnej teórie emócií.

Literatúra

- Atkinson, R. L., Atkinson, R. C., Smith, E. E., Bem, D. J. a Nolen-Hoeksema, S. (2003). *Psychologie*. Praha: Portál.
- Baron-Cohen, S. (2004). *Mind Reading Emotions Library* [CD]. Londýn: Jessica Kingsley Publishers.
- Basile, B., Mancini, M., Macaluso, E., Caltagirone, C., Frackowiak, R. S. a Bozzali, M. (2011). Deontological and Altruistic Guilt: Evidence for Distinct Neurobiological Substrates. *Human Brain Mapping*. 32(2):229–239.
- Beauregard, M., Lévesque, J. a Bourgouin, P. (2001). Neural Correlates of Conscious Self-Regulation of Emotion. *The Journal of Neuroscience*. 21(18):RC165.
- Bechara, A., Damasio, H. a Damasio, A. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Oxford Journals*. 10(3):295–307.
- Campbell, M. W. a de Waal, F. B. M. (2014). Chimpanzees Empathize with Group Mates and Humans, but not with Baboons or Unfamiliar Chimpanzees. *Proceedings of the Royal Society B*. 281:20140013.
- Ekman, P. (2003). *Emotions Revealed*. New York: Times Books.

- Ellsworth, P. C. a Scherer, K. R. (2003). Appraisal Processes in Emotion. V zborníku *Handbook of Affective Sciences*, str. 572–595.
- Fourie, M. M., Thomas, K. G. F., Amodio, D. M., Warton, C. M. R. a Meintjes, E. M. (2014). Neural Correlates of Experienced Moral Emotion: An fMRI Investigation of Emotion in Response to Prejudice Feedback. *Social Neuroscience*. 9(2):203–218.
- Fox, G. R., Kaplan, J., Damasio, H. a Damasio, A. (2015). Neural Correlates of Gratitude. *Frontiers in Psychology*. 6:1491.
- Gilead, M., Katzir, M., Eyal, T. a Liberman, N. (2015). Neural Correlates of Processing “Self-Conscious” vs. “Basic” Emotions. *Neuropsychologia*. 81:207–218.
- Goleman, D. (1997). *Emoční inteligence*. Praha: Columbus.
- Hammond, M. (2006). Evolutionary Theory and Emotions. V zborníku *Handbook of the Sociology of Emotions*, str. 368–385.
- Hill, G. (2004). *Moderní psychologie*. Praha: Portál.
- Izard, C. E. (1991). *The Psychology of Emotions*. New York: Plenum.
- Jurášek, D. a Marman, P. (2014). Kognitívna veda a vôľa. V zborníku *Kognice a umělý život XIV*, str. 97–100.
- Kahneman, D. (2013). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kassin, S. (2007). *Psychologie*. Brno: Computer Press.
- Kédia, G., Berthoz, S., Wessa, M., Hilton, D. a Martinot, J. L. (2008). An Agent Harms a Victim: A Functional Magnetic Resonance Imaging Study on Specific Moral Emotions. *Journal of Cognitive Neuroscience*. 20(10):1788–1798.
- Keltner, D., Ekman, P., Gonzaga, G. C. a Beer, J. (2003). Facial Expression of Emotion. V zborníku *Handbook of affective sciences*, str. 415–432.
- Kubáni, V. (2010). *Všeobecná psychológia*. Prešov: Prešovská univerzita v Prešove.
- Leary, M. R. (2007). Motivational and Emotional Aspects of the Self. *Annual Review of Psychology*. 58:22.01–22.28.
- Lindquist, K. A., Wager, D. T., Kober, H., Bliss-Moreau, E. a Barret, L. F. (2012). The Brain Basis of Emotion: A Meta-Analytic Review. *Behavioral and Brain Sciences*. 35(3):121–143.
- Nakonečný, M. (2012). *Emoce*. Praha: Triton.
- Phan, K. L., Wager, T., Taylor, S. F. a Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage*. 16:331–348.
- Plutchik, R. (2002). *Emotions and Life: Perspectives from Psychology, Biology and Evolution*. Washington: American Psychological Association.
- Reimerta, I., Bolhuis, J. E., Kemp, B. a Rodenburg, T. B. (2013). Indicator of Positive and Negative Emotions and Emotional Contagion in Pigs. *Physiology and Behavior*. 109:42–50.
- Robinson, D. L. (1998). Brain function, emotional experience and personality. *Netherlands Journal of Psychology*. 64:152–167.
- Roth, L., Kaffenberger, T., Herwig, U. a Brühl, A. B. (2014). Brain Activation Associated with Pride and Shame. *Neuropsychobiology*. 69(2):95–106.
- Rutherford, K. M., Donald, R. D., Lawrence, A. B. a Wemelsfelder, F. (2012). Qualitative Behavioural Assessment of Emotionality in Pigs. *Applied Animal Behaviour Science*. 139(3-4):218–224.
- Sternberg, R. J. (2002). *Kognitivní psychologie*. Praha: Portál.
- Stuchlíková, I. (2002). *Základy psychologie emocí*. Praha: Portál.
- Tracy, J. L. a Robins, R. W. (2004). Putting the Self into Self-Conscious Emotion: A Theoretical Model. *Psychological Inquiry*. 15:103–125.
- Vytal, K. a Hamann, S. (2010). Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-Analysis. *Journal of Cognitive Neuroscience*. 22(12):2864–2885.
- Warneken, F., Hare, B., Melis, A. P., Hanus, D. a Tomasello, M. (2007). Spontaneous Altruism by Chimpanzees and Young Children. *PLoS Biol.* 5(7):e184.

Dôvera a nedôvera vo vedecké a nevedecké informácie

Marek Jurkovič

Centrum spoločenských a psychologických vied SAV: Ústav experimentálnej psychológie

Dúbravská cesta 9, 841 04 Bratislava, Slovenská Republika

E-mail: marek.jurkovic@savba.sk

Abstrakt

Pre súčasného človeka je dnes prístup k informáciám po technickej stránke veľmi jednoduchý a s nástupom sociálnych médií sa aj podieľajú na ich tvorbe a masovom šírení. Ako však ľudia filtrujú podložené informácie od nepodložených? Na základe čoho sa rozhodujú, že určitým informáciám uveria alebo neuveria? Vedecké poznatky sa stávajú terčom nedôvery alebo sú skresľované, kým nepodložené fámy sú nekriticke šírené ďalej. V tejto teoretickej štúdiu som sa zamerlal na zhromaždenie poznatkov, ktoré by mohli poskytnúť určitý obraz o príčinách tohto stavu z psychologickéj perspektívy a možnostiach ďalšieho výskumu v tejto oblasti. Motiváciou pre tento článok je snaha upozorniť, že nedôvera vo vedu môže mať negatívne následky pre vedcov aj nevedcov.

1 Úvod

V súčasnosti sa vede venuje v bežných médiách výrazne menej priestoru, než iným oblastiam, napriek neustávajúcemu prílevu nových objavov, ktoré môžu ovplyvniť naše životy a spoločnosť. Zároveň možno pozorovať, že spoločne s väčšou dostupnosťou informácií prostredníctvom internetu majú ľudia na dosah aj veľké množstvo dezinformácií, čo sa podieľa na fenoménoch ako napr. "Doktor Google", keď jednotlivci na základe informácií získaných z internetu dospievajú k názoru, že im lekári taja zaručene účinné lieky na rakovinu, prípadne, že si vedia utvoriť lepšie informovaný názor na problematiku klimatických zmien, evolučnú teóriu alebo národnú históriu (či akúkoľvek inú problematiku), než sú toho schopní odborníci v danej oblasti. Čo sa z psychologického hľadiska skrýva za týmto fenoménom a čo sa s tým dá robiť?

1.1 Porozumenie vedeckým poznatkom a ich akceptácia

Dalo by sa predpokladať, že neakceptovanie vedeckých zistení môže mať pôvod v nedostatočných vedomostiach jednotlivcov v tejto oblasti. Ak vychádzame zo zistení Krugera a Dunninga (1999), je možné, že jednotlivci s malými poznatkami v určitých

kontroverzných oblastiach by mohli preceňovať svoju schopnosť posudzovať správnosť vedeckých zistení. Takáto predstava je možno v jednotlivcoch posilňovaná možno aj povrchným mediálnym referovaním o komplexných vedeckých výsledkoch, nakoľko ani žurnalisti venujúci sa vede majú málokedy možnosť alebo motiváciu byť odborníkmi v odboroch o ktorých podávajú informácie. V takom prípade by sme mali zistiť, že spoločne s hlbším porozumením vedeckej teórie nastane aj jej akceptácia. Prípadne môžeme predpokladať opačnú pozíciu, že na porozumenie vedeckým teóriám je nutné najprv adresovať ich akceptáciu "poslucháčmi", inak poskytované informácie budú jednoducho ignorované (Smith, 2007). Ak si však zoberieme príklad evolučnej teórie, na základe dostupnej literatúry zistíme, že akceptácia nie je spojená s porozumením teórii (Bishop & Anderson, 1990; Demastes, Settlege, & Good, 1995). Aj napriek nárastu porozumenia evolučnej teórii sa akceptácia evolučnej teórie v oboch citovaných štúdiách významne nezmenila.

Podobné výsledky možno nájsť aj v iných vedomostných doménach, napríklad v doméne akceptácie antropogénnych klimatických zmien (Kahan, 2015).

Čo bráni jednotlivcom v tom, aby akceptovali teórie, ktorým rozumejú?

1.2 Princípy vedy a akceptácia vedeckých poznatkov

Čiastočne je to možno dané chýbajúcim porozumením princípov vedy. Ako zistili Lombrozo, Thanukos a Weisberg (2008), akceptácia evolučnej teórie bola významne spojená s porozumením vedeckých princípov. Pod povahou vedy sa myslia napríklad také koncepty, ako provizionalita vedeckých teórií a ich neprestajné overovanie pomocou nových pozorovaní a experimentov. Čiastočne sa to môže prejavovať tým, že sa vďaka porozumeniu vedeckých princípov stávajú pseudoargumenty typu "je to len teória" menej presvedčivé, nakoľko jednotlivec už je schopný spoznať vadu v takomto pseudoargumente. K nesprávnemu pochopeniu princípov vedy pravdepodobne vedie aj pomerne nesprávne vysvetľovanie v školách, kde testovanie hypotéz je

vysvetľované v postup vytvorenia laboratórneho experimentu, ktorý hypotézu podporí alebo nepodporí (v horšom prípade dokonca "dokáže" alebo "vyvráti"), pritom takýto postup nereflektuje rozdielne vedecké postupy napr. aj v rámci evolučnej biológie. Navrhovanou možnosťou nápravy je dať študentom aspoň zjednodušené vedomosti týkajúce sa princípov vedy, ktoré by však zároveň zachytávali väčšiu komplexnosť, než učebnicové vysvetlenia vedeckej metódy (Lombrozo et al., 2008). Je otázne ako by sa o týchto princípoch dali informovať jednotlivci, ktorí už štúdium ukončili.

Aj v tomto prípade (Lombrozo et al., 2008) však výskumníci zistili, že porozumenie princípom vedy má efekt len na jednotlivcov, ktorí ešte nie sú presvedčení o pravdivosti alternatívnych vysvetlení (v prípade evolučnej teórie sa jednalo o presvedčenie, že život na Zemi bol stvorený a nevyvíja sa - kreacionizmus resp. "inteligentný dizajn").

Existuje teda skupina jednotlivcov, ktorí vedeckým teóriám rozumejú a prípadne aj disponujú poznatkami týkajúcimi sa princípov vedy. Napriek tomu zotrávajú v presvedčeniach, ktoré nie sú podporené vedeckým konsenzom, podložené empirickými dátami alebo jednoducho nespĺňajú základné požiadavky vedeckej teórie. Čo vedie jednotlivcov k odmietaniu vedeckých zistení v takýchto prípadoch?

2 Akceptácia vedeckých zistení a identita

Jedno zo zistení v už citovaných štúdiách (Bishop & Anderson, 1990; Demastes et al., 1995; Kahan, 2015; Lombrozo et al., 2008; Smith, 2007) je, že participanti akceptovali resp. neakceptovali vedecké poznatky na základe náboženského presvedčenia (v prípade evolučnej teórie) alebo politickej afiliácie (v prípade klimatických zmien). Minimálne u konzervatívnej časti populácie v USA sa tiež zistila klesajúca dôvera (v priebehu rokov 1974 až 2010) vo vedu a to aj u konzervatívov s vyšším vzdelaním (Gauchat, 2012).

Ako to vysvetľuje Kahan (2015), ide o rozpor medzi tým, čo človek vie a čím je. Kahan (2015) zistil, že hoci môžu participanti odpovedať správne aj na značne zložité otázky týkajúce sa vedeckej gramotnosti (ktorú možno chápať ako vedomosti týkajúce sa uznávaných vedeckých poznatkov), v prípade súhlasu alebo nesúhlasu s tvrdením, že sa človek evolúciou vyvinul z iných druhov je vysoká pravdepodobnosť, že budú odpovedať na základe toho, ako sa kultúrne identifikujú na základe náboženstva, nie na základe vedecky uznávaných faktov. Rozdielna formulácia otázky viedla k odlišným výsledkom, keď sa otázka pýtala na to, či podľa teórie evolúcie je vývin človeka z predošlých druhov pravdivý alebo nepravdivý. Identita teda nebráni porozumeniu vedeckých poznatkov, bráni len ich akceptácii. Podobné výsledky zistil (Kahan, 2015) aj v prípade politickej afiliácie a posudzovania

pravdivosti/nepravdivosti klimatických zmien následkom ľudskej činnosti.

Jedným z podstatných zistení týkajúcich sa vedeckej gramotnosti vo vzťahu ku kontroverzným a k identite relevantným témam je zistenie, že vedecká gramotnosť v takýchto prípadoch nemusí jednotlivcov chrániť pred neakceptovaním uznávaných vedeckých poznatkov (Kahan et al., 2012).

V rámci výskumu racionality a kognitívnych zaujatostí existuje predpoklad, že kognitívne zaujatosti sú do istej miery spôsobované heuristickým uvažovaním, pričom za racionálnejšie sa považuje pomalšie analytické myslenie, ktoré tieto heuristické reakcie potláča. U jednotlivcov s vyššou úrovňou analytického myslenia by sme teda mali pozorovať menšie podliehanie rôznym zaujatostiam podmieňujúcim iracionalitu a tým pádom aj vyššiu akceptáciu podložených vedeckých teórií, pretože jednotlivci by mali byť schopní rozumieť zložitejším modelom a rozumieť tomu, prečo sú správne (Weber & Stern, 2011).

Lenže práve v prípade kontroverzných a polarizujúcich tém pozorujeme opak. Hoci jednotlivci disponujúci vysokou úrovňou vedeckej gramotnosti a mali by byť teda schopní porozumieť aj komplexným modelom a disponujú rozvinutým analytickým myslením, miera akceptácie teórií je u nich najviac skreslená ich identitou (Kahan et al., 2012).

To neznamená, že by vedecká gramotnosť alebo analytické myslenie nesúviseli s mierou akceptácie vedeckých zistení, no tento vzťah sa prejavuje pri nekontroverzných nepolarizujúcich témach (Kahan, 2015). Zvyšovanie vedeckej gramotnosti a úrovne analytického myslenia je adekvátny cieľ, no nemusí sa prejaviť vtedy, keď je postoj k tvrdeniu súčasťou identity jednotlivca.

Toto zistenie možno dať do súvislosti s tzv. argumentatívnou teóriou. Podľa tejto teórie neslúži uvažovanie na získavanie a vylepšovanie vedomostí, či robenie lepších rozhodnutí. Uvažovanie podľa tejto teórie slúži na presvedčanie a podporovanie vlastných presvedčení. Pomocou tejto teórie možno vysvetliť aj napríklad konfirmačnú zaujatosť, teda tendenciu vyhľadávať a vyhodnocovať informácie podľa toho, či podporujú naše presvedčenia (Mercier & Sperber, 2011). To je umožnené aj tým, že realita síce do istej miery obmedzuje naše presvedčenia (v stave relatívneho duševného zdravia nemožno popierať úplne zrejme skutočnosti), no napriek tomu naše chápanie sveta a informácie o ňom sú často značne nejednoznačné a často existuje viacero alternatívnych vysvetlení jedného fenoménu, čo umožňuje voľnosť vo vytváraní presvedčení a v skresľovaní informácií v prospech našich presvedčení (Gilovich, 1993).

Ako možno zabrániť prejavovaniu konfirmačnej zaujatosti resp. akceptácii informácií vzhľadom na určitú kultúrnu identitu? Kahan (2015) navrhuje formulovať kontext polarizujúcich tém tak, aby nevyvolávali tendenciu obhajovať vlastnú identitu. To znamená, že by sme sa mali pri riešení problémov ako

klimatické zmeny mali obracať skôr na to, čo ľudia vedia o danom probléme a ako by ho mohli riešiť, nie na to, kým sú, nevyvolávať polarizáciu na základe príslušnosti k určitej skupine pomocou nesprávneho komunikovania informácií, napr. práve podceňovaním určitej skupiny ľudí a ich schopnosti pracovať s vedeckými informáciami.

2.1 Identita a komnaty ozveny

S konfirmačnou zaujatou a posudzovaním informácií na základe identity súvisí aj fenomén "komnaty ozveny". Je to fenomén, keď sa v rámci komunity posilňujú a opakujú určité presvedčenia a protichodné presvedčenia sú cenzurované, potláčané alebo sú reprezentované menej.

Jedným z príkladov je vytváranie komunit na internete a sociálnych sieťach. V súčasnosti sú rôzne dezinformácie rozširované po internete takmer bez prekážok a vznikajú komunity, ktoré sa zameriavajú na šírenie a konzumáciu "alternatívnych" informácií, ktoré sú veľmi často v protiklade s vedeckými vysvetleniami. Inak povedané sa jedná o konzumentov rôznych konšpiračných teórií (KT), fám, hoaxov a všeliakov. Akceptácia konšpiračných teórií je tiež spojená s odmietaním vedeckých zistení (Lewandowsky, Gignac, & Oberauer, 2013). Viera v KT sa môže tiež stať súčasťou identity (Raab, Ortlieb, Auer, Guthmann, & Carbon, 2013) a slúžiť ako istý druh neorganizovaného náboženstva slúžiaceho na vysvetľovanie skutočnosti (Franks, Bangerter, & Bauer, 2013). Existujú teda predpoklady pre polarizáciu presvedčení na základe príslušnosti ku komunite ľudí presvedčených o pravdivosti KT.

Spoločne s vytváraním internetových komunit zameraných na konzumáciu istého obsahu sa však prejavuje aj selektívny výber informácií, teda efekt komnaty ozveny. Tento efekt sa neprejavuje len u vedecky nepodložených presvedčení, akými sú rôzne KT. U konzumentov vedeckých informácií tiež možno pozorovať sklony k preferovaniu jedného druhu informácií, podobne ako u konzumentov KT (Bessi et al., 2015). Problematické pri KT je však to, že konzumenti KT sa viac zameriavajú len na šírenie a akceptáciu informácií v porovnaní s konzumentmi vedeckých informácií, ktorých online činnosť je do väčšej miery spojená aj s konfrontáciou s odlišnými názormi (teda práve pseudovedeckými informáciami). Konzumenti KT sa teda viac uzatvárajú do komnaty ozveny. U oboch komunit sú ale informácie preberané prevažne od priateľov s rovnakými záujmami alebo presvedčeniami, teda od priateľov patriacich k rovnakej komnate ozveny (Del Vicario et al., 2016).

Tieto zistenia možno dať do súvislosti aj s personalizáciou výsledkov vyhľadávania v internetových vyhľadávačoch, či s personalizáciou zobrazovaných komunit a stránok na sociálnych sieťach. Personalizácia týmto tiež pravdepodobne

prispieva ku konfirmačnej zaujatosti a k vytváraniu komnat ozvien.

Je otázne, ako by sa táto tendencia vytvárať názorovo jednostranné komunity dala narušiť v prospech vyhľadávania odlišných informácií a názorov, najmä u jednotlivcov, ktorí veria v KT a rôzne pseudovedecké nepodložené tvrdenia. Jednou z problematických charakteristík KT je ale aj interpretácia snáh o ich vyvrátenie ako súčasť konšpirácie, čo z KT robí presvedčenia odolné voči zmene (Lewandowsky et al., 2013).

Jednou z možných stratégií je tiež zvyšovanie dôvery vo vedecké zistenia prostredníctvom zvyšovania kvality výskumu. Snahy o zvyšovanie dôvery v odborné autority by ale v takom prípade boli stále podkopávané rôznymi záujmovými skupinami, ktorých finančná situácia alebo popularita závisí od šírenia nepodložených informácií. Opäť sa ale tiež možno snažiť komunikovať vedu spôsobom, ktorý nevyvoláva polarizáciu na základe príslušnosti k skupine a obrany kultúrnej identity.

3 Veda o komunikácii vedy

Komunikovanie vedeckých poznatkov zdá sa stále na naivných predstavách, že stačí v populácii zvyšovať informovanosť o problematike alebo zvyšovať mieru analytického myslenia. Ako možno však využiť z predošlého textu, tento predpoklad je síce v zásade pravdivý v nekontroverzných kontextoch, ale práve v kontextoch pri ktorých sa prejavuje najvyššia polarizácia spoločnosti sa pozitívny efekt vzdelania a kognitívnej kapacity obracia proti jednotlivcovi. Pri komunikácii vedeckých poznatkov populácii tiež chýbajú vedecké poznatky, ktoré by sa týkali účinnej komunikácie vedy, ktorá by nepodporovala v spoločnosti konflikty (Kahan, 2014) a ktorá by dokázala premostiť problém konzumácie jednostranných informácií. Jedným z príkladov neúčinnnej komunikácie vedy a snahy pozitívne zapojiť ľudí je vyvolávanie strachu v súvislosti s klimatickými zmenami (O'Neill & Nicholson-Cole, 2009). Určité návrhy už boli spomenuté, no ich realizácia v praxi zatiaľ nie je podporená empiricky.

Van der Linden, Maibach a Leiserowitz (2015) navrhujú týchto 5 prístupov pri komunikácii poznatkov súvisiacich s klimatickými zmenami:

- Zdôrazňovanie osobných skúseností, teda konkrétne príklady ako klimatické zmeny škodia ovplyvňujú životné prostredie
- Využívanie relevantných sociálnych noriem
- Zdôrazňovať lokálne dopady klimatických zmien z nedávnej minulosti, teda znižovať psychologickú vzdialenosť
- Framing možných opatrení v rámci toho, čo sa nimi dá získať a nie, čo bude stratené
- Využívať ľudskú motiváciu (napr. budovanie lepšej budúcnosti)

Viacere tieto pristupy nesluzia ani tak na akceptáciu zistení, ako skor na vzbudenie aktivity u jednotlivcov. Adaptácia na lokálne podmienky by však mohla zabrániť efektu kultúrne ovplyvnenej kognície, pretože viac než o vojnu dvoch názorových prúdov sa v lokálnom kontexte jedná napr. o ochranu vlastného majetku a pod. (napr. stúpajúcou hladinou mora) (Kahan, 2014). Budúce výskumy by sa mali tiež realizovať v snahe o čo najväčšiu ekologickú validitu a mali by zohľadňovať lokálny kontext, laboratórne výskumy totiž nemusia spustiť sociálne vplyvy ako napr. efekt komnaty ozveny alebo kultúrnej identity. Veda o komunikácii vedy je výzvou pre budúci výskum, ktorého spoločenskú dôležitosť nemožno poprieť vzhľadom na množstvo v súčasnosti šírených dezinformácií a polarizáciu v konfliktných témach.

Pod'akovanie

Príspevok bol podporený grantovou agentúrou VEGA 2/0064/13: Rozhodovanie expertov: využívanie intuície expertmi pri riešení strategických úloh.

Literatúra

- Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociochi, W. (2015). Science vs Conspiracy: Collective Narratives in the Age of Misinformation. *Plos One*, 10(2), e0118093. doi:10.1371/journal.pone.0118093
- Bishop, B. A., & Anderson, C. W. (1990). Student Conceptions of Natural Selection and Its Role in Evolution *. *27(5)*, 415–427.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociochi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 201517441. doi:10.1073/pnas.1517441113
- Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching*, 32(5), 535–550. doi:10.1002/tea.3660320509
- Franks, B., Bangerter, A., & Bauer, M. W. (2013). Conspiracy theories as quasi-religious mentality: An integrated account from cognitive science, social representations theory, and frame theory. *Frontiers in Psychology*, 4(JUL), 1–13. doi:10.3389/fpsyg.2013.00424
- Gauchat, G. (2012). Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010. *American Sociological*

Review, 77(2), 167–187. doi:10.1177/0003122412438225

- Gilovich, T. (1993). *How We Know What Isn't So*. Free Press. Retrieved from <https://books.google.sk/books?id=tHhKSLoORqIC>
- Kahan, D. M. (2014). Making Climate-Science Communication Evidence-based—All the Way Down. *Culture, Politics and Climate Change*, 203–221. doi:10.2139/ssrn.2216469
- Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, 36(S1), 1–43. doi:10.1111/pops.12244
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10), 732–735. doi:10.1038/nclimate1547
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–34. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10626367>
- Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *PLoS ONE*, 8(10). doi:10.1371/journal.pone.0075637
- Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The Importance of Understanding the Nature of Science for Accepting Evolution. *Evolution: Education and Outreach*, 1(3), 290–298. doi:10.1007/s12052-008-0061-8
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *The Behavioral and Brain Sciences*, 34(2), 57–74; discussion 74–111. doi:10.1017/S0140525X10000968
- O'Neill, S., & Nicholson-Cole, S. (2009). "Fear Won't Do It": Promoting Positive Engagement With Climate Change Through Visual and Iconic Representations. *Science Communication*, 30(3), 355–379. doi:10.1177/1075547008329201
- Raab, M. H., Ortlieb, S. a., Auer, N., Guthmann, K., & Carbon, C. C. (2013). Thirty shades of truth: Conspiracy theories as stories of individuation, not of pathological delusion. *Frontiers in Psychology*, 4(JUL), 1–10. doi:10.3389/fpsyg.2013.00406

Smith, M. U. (2007). Counterpoint Belief, understanding, and the teaching of evolution. *Journal of Research in Science Teaching*, 31(5), 591–597. doi:10.1002/tea.3660310512

van der Linden, S., Maibach, E., & Leiserowitz, A. (2015). Improving Public Engagement With Climate Change: Five “Best Practice” Insights From Psychological Science. *Perspectives on Psychological Science*, 10(6), 758–763. doi:10.1177/1745691615598516

Weber, E. U., & Stern, P. C. (2011). Public understanding of climate change in the United States. *American Psychologist*, 66(4), 315–328. doi:10.1037/a0023253

Když je více méně: Význam lidské chyby při interakci s 3D virtuálním geografickým prostředím

Juřík V., Šašík Č.

Centrum experimentální psychologie a kognitivních věd, Psychologický ústav FF MU, Arna Nováka 1, Brno
372092@mail.muni.cz, 44276@mail.muni.cz,

Abstrakt

Vizualizace v oblasti interakce člověk-počítač vstoupily do éry 3D. Virtuální reprezentace prostorových dat jsou stále častěji využívány v aplikovaných oblastech od vojenství, navigace, přes krizové řízení až po leteckou kontrolu (ATC). Přidaná hodnota trojdimenzionálních reprezentací spočívá v objektivně větším množství vizuálních nápovědí, které při práci s nimi může člověk využít. Na druhé straně, z psychologické perspektivy vyvstává otázka, do jaké míry daný typ vizualizace mění povahu vnímaného, a to nejen na percepční, ale také na interakční úrovni. Naše předchozí studie ukazují, že navzdory informační ekvivalenci prezentovaného obsahu, forma 3D vizualizace ovlivňuje výkon participantů. Jako zásadní vnímáme zjištění, že imerzivní reálná 3D vizualizace, navzdory většímu množství vizuálních nápovědí a realističnosti, může u člověka zvýšit tendenci opomínat významné aspekty prezentované expozice, což v důsledku vede k nezvládnutí zadaného úkolu. V oblasti interakce člověk-počítač, kde může relativně malá chyba stát lidské životy, je např. zvýšení rychlosti procesu vyhodnocení dat za cenu nepřepokládané chyby zásadní otázka. Primárně by se ve vývoji uživatelských rozhraní nemělo jednat o snížení chybovosti, ale o vyloučení jakékoli tendence k chybování. Pro tento účel je nezbytné podrobně analyzovat okolnosti systému stejně jako proces, jakým operátor do interakce se systémem vstupuje. V závislosti na výsledcích našeho dosavadního výzkumu navrhujeme uživatelské rozhraní, které zajišťuje otevřené možnosti interakce člověk-počítač a zároveň umožňuje jejich efektivní měření. Za použití moderních technologií (Eye-tracking systém, Motion Capture systém, širokouhlá zpětná 3D projekce) je možné analyzovat strategie řešení problému v různých typech 3D virtuálních rozhraní.

1 3D zobrazení ve VR

Virtuální vizualizace vstoupily do éry 3D (Boughzala, 2012). Pro dosažení lepší věrohodnosti virtuálních prostředí, tzv. fidelity (Hochmitz & Yuviler-Gavish, 2011), jsou v praxi často využívány 3D technologie, které dokáží věrně simulovat prostorové charakteristiky reálného prostředí. Mapy, coby zásadní grafické reprezentace prostorového uspořádání, jsou vzhledem k frekvenci svého využívání transformovány do 3D verzí s ohledem na obsah, který má být prostřednictvím mapy komunikován. Uživatelská přívětivost 3D vizualizací je podrobně diskutovaná problematika (Schmidt, Delazari, Mendonca, 2012). V oblastech, jakými jsou letectví, krizové řízení, přírodní hazardy a nebo vojenství (Maggi et al., 2015, Hirmas et al., 2014; Wilkening & Fabrikant, 2013), kde bývá základní prostorová informace obvykle znázorněna typem kartografické vizualizace, se vhodnost uživatelského rozhraní jeví jako klíčová výzkumná otázka. Daný typ 3D vizualizace může ovlivnit povahu proximálních nápovědí, které jedinec vybírá a uvažuje při řešení úkolu (Hammond & Stewart, 2001). Pro pochopení vlastností virtuálních prostředí a jejich vlivu na lidskou kognici je nezbytné analyzovat senzomotorické principy percepce, které determinují člověka při procesu vnímání a následné interpretaci vnímaného. Předpokládáme, že typy proximálních nápovědí, které operátor při práci s virtuálním prostředím považuje za relevantní, a dále je percipuje a interpretuje, ovlivňují kognitivní procesy stejně jako konečné výstupy celé interakce. Povaha prostředí reprezentovaná typem vizualizace může představovat zásadní determinant lidského faktoru v oblasti interakce člověk-počítač. Nehody obvykle nejsou způsobeny jediným faktorem, ale souhrou technických, systemických a lidských faktorů. Primárně je to ale lidský faktor, který bývá považován jako hlavní viník nehod (Woods et al., 2010). Průzkumy incidentů přisuzují 70 % všech

kritických událostí jako zapříčiněných lidskou chybou (Amalberti, 2001; Hollywell, 1996).

Jak poznamenávají někteří výzkumníci (Kontogiannis & Malakis, 2009; Amalberti, 2001), úplné eliminace lidské chyby v systémech není možné dosáhnout. V oblasti virtuálních prostředí, která jsou vyvíjena mimo jiné pro simulaci a výzkum rizikových situací, je z psychologického hlediska nezbytné zvážit roli anticipačních schémat, která v konsonanci s Neisserovým percepčním cyklem (1967) mohou poskytnout náhled na problematiku interakce člověk-počítač, jmenovitě v oblasti lidské chyby (human error). Jak poznamenává Plant a Stanton (2012), Neisserův percepční model je možné využít při snaze o pochopení lidské chyby, a tudíž také v následném vývoji systémů s důrazem na prevenci lidské chyby.

2 Diskutované aspekty virtuální reality

Virtuální realita (VR) je některými výzkumníky vnímána jako využití technologie pro vytvoření 3D interaktivního a imerzivního světa, kde se uživatel cítí být zanořen a přítomen (Bryson, 1999). Počítačem generovaná virtuální prostředí jsou tedy slibnou oblastí nejen z hlediska snadné modifikace, aktualizace a přizpůsobivosti obsahu, ale také díky přesvědčivosti daného prostředí, které efektivně simuluje podmínky reálného světa. Vysoce imerzivní virtuální světy umožňují fyzicky vzdáleným operátorům zastoupeným jejich avatary interagovat a efektivně spolupracovat (Davis, Murphy, Owens, Khazanchi, & Zigurs, 2009). Během procesu takové kooperace mohou operátoři komunikovat a koordinovat práci a úkoly (Boughzala, 2012). Z tohoto hlediska jsou možnosti umělých 3D prostředí bohaté a jejich aplikační potenciál je značný. Virtuální světy zajišťují kontrolovatelné a dobře nastavitelné prostředí pro psychologický výzkum kognitivních procesů, kde je možné zajistit ekologicky validní stimuly a přesné záznamy behaviorální aktivity probandů (Wilson & Soranzo, 2015). Domníváme se, že každý typ umělého rozhraní by měl zohledňovat účel, ke kterému je vyvíjen, protože rozdílné formy vizualizace ovlivňují různé fenomény vstupující do procesu interakce. Rozdílné parametry designu rozhraní mohou vyvolat odlišnou psychickou odezvu člověka. Tyto je nezbytné reflektovat v rámci navrhování prostředí. Diskutováním pojmem, které se pojí s vývojem a uživatelskými aspekty virtuálních rozhraní, je fenomén presence (Mania et al., 2006; Björk & Holopainen, 2004; Mania & Chalmers, 2001; Slater & Wilbur, 1997), někdy se označuje také jako tzv. vnoření do VR (immersion). Aspekt presence je považován za pozitivní efektor při vztahu mediální synchronicity a spokojenosti s výsledkem činnosti (Hassell & Limayem, 2010), což jsou faktory zásadně ovlivňující motivaci a výkon při řešení úkolů ve VR. Problematika učení ve VR bývá často spojována právě

s navozením dojmu presence (Bailenson a kol., 2006). Dále stojí ve středu výzkumného zájmu sociální interakce ve 3D VR (Goel & Prokopec, 2009). Dalšími oblastmi jsou situační povědomí (Endsley, 1995), kognitivní zátěž nebo lidské pochybení (Zimmerman & Koebbe, 2013; Nunez, 2005; Tavanti et al., 2003). Zmiňované koncepty jsou uváděny do souvztažnosti, kde zvyšující se kognitivní zátěž, ve snaze nabídnout maximální množství informace (např. zobrazení příliš velkého množství informace), zvyšuje riziko chybování (Forsell, 2007; Lange, 2003). Z tohoto úhlu pohledu diskutuje Hammond a Stewart (2001) problematiku proximálních náповědí obsažených v typu vizualizace (potažmo rozhraní) a jejich vlivu na rozhodování člověka-operátora. Zároveň může mít velké vnoření do VR za následek opomenutí konkrétního aspektu scény, což může vést k chybě. Domníváme se, že i v informačně ekvivalentních (Larkin & Simon, 1987) prostředích mohou být na základě rozdílného typu vizualizace percepčně upřednostňovány specifické proximální náповědi, které ovlivní proces interakce s vizualizací a tudíž také evaluaci a interpretaci viděného. Otázka lidského faktoru, jmenovitě fenomén lidské chyby, je v rámci interakce-člověk počítač ústřední téma vzhledem k tomu, že často přímo způsobuje ohrožení člověka nebo majetku. Nowakowski a kolegové (2015) zdůrazňují, že lidská chyba je nezamýšlenou akcí, která překračuje limity daného systému a s využitím ergonomického designu je možné vytvořit systém eliminující pravděpodobnost lidského omylu (Hankiewicz, 2012). Proto je důležité zkoumat senzomotorické a kognitivní procesy, které do interakce vstupují. Lidská chyba v rámci systému nemusí nutně vést k nehodě samotné, nelze ji tedy považovat jako kauzální prediktor selhání systému. Konsekvence z ní vyplývající mohou nabývat různé formy. Reason (1990) uvádí, že lidská chyba může být způsobena i v dobré víře, např. několika rozhodnutími, která byla adekvátní v daném čase, ale nakonec vedla k selhání systému. Proto je nezbytné zkoumat širší souvislosti, které k chybě vedly.

3 Lidská chyba v oblasti 3D vizualizací

Z metodologického hlediska není možné nahlížet na lidskou chybu čistě z kvantitativního hlediska, někteří výzkumníci (Dekker, 2002) vnímají lidské pochybení spíše jako abstraktní označení problematiky než fenomén, který lze vyčíslit. Lipshitz a kolegové (2001) tvrdí, že by měly být studovány a popisovány především faktory ovlivňující okolnosti situace než pouze kvantifikovat četnost chyb. Lidská chyba z kvantitativního úhlu pohledu tvoří minoritní fenomén, který nastává s relativně malou pravděpodobností, nicméně je spojován s extrémními náklady a možnými ztrátami na životech (Stanton & Walker, 2011). Toto je nezbytné zohlednit při tvorbě výzkumného designu, kde se nelze spoléhat na standardní statistické postupy.

Nezbytné je zkoumat kontext prostředí a anticipovat parametry prostředí, které mohou k lidské chybě vést, což vyzývá ke zkoumání spíše kvalitativní povahy. V oblasti lidské chyby je vždy nezbytné zohledňovat kontext situace (Bennett, 2001; Dekker, 2006; Woods et al., 2010). Cacciabue (2004) zdůrazňuje, že předmětem metodologie zkoumání lidské chyby by neměly být pouze kvantitativní rozdíly v úspěšnosti resp. neúspěchu v konkrétní úloze, ale hledání pozadí, na kterém se chyba zrodila.

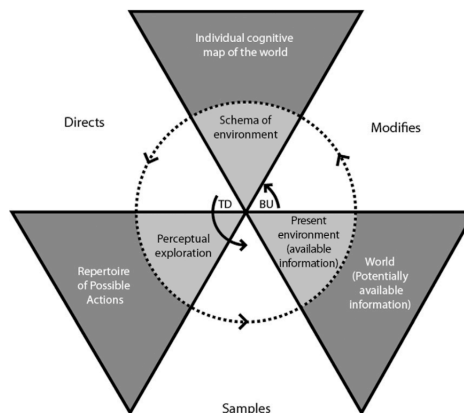
Lidské selhání je trvale považováno za hlavní příčinu nehod v rizikových oblastech, jako je například letectví (Stanton & Salmon, 2009; Amalberti, 2001). Přes výraznou automatizaci systémů je to vždy operátor, který vyhodnocuje situaci a je proto nezbytné zkoumat a porozumět jeho kognitivním procesům, které vedly k případnému selhání (Plant & Stanton, 2012). V současné době se prosazuje snaha o postihnutí procesu lidského selhání v širším kontextu situace, o porozumění okolnostem a procesům, které k lidské chybě vedly (Dekker, 2002; Reason, 1990). Fedota and Parasuraman (2009) argumentují, že je třeba zohledňovat jak starý přístup k problematice human factors – tedy že hlavním viníkem selhání je vždy operátor, a zároveň také nový model, který hledá vysvětlení na straně okolností a prostředí. Plant a Stanton (2012) shledávají pro tuto syntézu jako adekvátní model percepčního cyklu (Neisser, 1967), který s oběma konci této problematiky pracuje (viz. Obr. 1). Dále zmiňují anticipační schémata, která vstupují do procesu percepce.

3.1 Model percepčního cyklu

Neisser (1967) již před více než 40 lety vyjádřil názor, že myšlení je úzce spojené s lidskou interakcí s prostředím, kde se jedná o oboustranný cyklický proces vzájemného ovlivňování – člověk ovlivňuje prostředí a prostředí zpětně ovlivňuje parametry lidské percepce. Lidské znalosti o světě, které považuje za tzv. schémata, determinují člověka k předjímání specifického typu informace z prostředí. Tato anticipace určuje lidské chování, které vede k vyhledávání a vnímání konkrétních prvků v prostředí, stejně tak jako k dalšímu chápání a vyhodnocení informací. Schémata jsou na základě zkušenosti ustanovována, modifikována a aktualizována a potom zpětně ovlivňují to, jak bude člověk interagovat s prostředím. Neisserova práce ovlivnila náhled na kognici, která začala být vnímána jako „na aktivitě člověka založený proces“. Člověk přestal být v procesu kognice vnímán jako pasivní činitel, ale jako aktivně zapojený do průběhu události (Klein, 2008). Důraz ve výzkumu by tedy měl být kladen na vstupy a procesy spíše než na výstupy situací.

Plant a Stanton (2012) nalézají podobnost mezi Dekkerovým chápáním procesu lidské chyby a

Neisserovým percepčním cyklem. Dekkerův pohled na lidskou chybu (2006) zdůrazňuje analýzu situace a kontextu prostředí při studiu lidské chyby a také to, že v případě lidského jednání se nejedná o statický, ale dynamický proces, kde lidé mění svým chováním situaci, ale zároveň situace modifikuje lidské vnímání a chování, což je v zásadě model percepčního cyklu. Z tohoto úhlu pohledu je lidská chyba produktem dynamické interakce člověka a prostředí, kde se série lidských rozhodnutí mohly v daném čase jevit jako relevantní, ale postupně vedly k chybě (Plant & Stanton, 2012).



Obr. 1. Neisserův percepční cyklus, dle Plant & Stanton (2012)

3.2 Anticipační schémata

V obecnosti lze schémata chápat jako struktury, na základě kterých chápeme objektivní realitu a také to, jak realitu reprezentujeme v mentální rovině. Schémata rovněž pomáhají vést další chování organismu v prostředí. Poskytují instrukce naší kognici a organizují percipované informace (Chalmers, 2003). Schémata jsou tradičně vnímána jako smysluplně organizované struktury, které obsahují podschémat, jsou modifikována příchozími informacemi, mění se v krátkých cyklech a nabývají celostní povahy (Anderson, 1977). Neisser (1976) definoval genotypní a fenotypní schémata. Genotypní schémata vnímá jako struktury, které vedou aktivitu organismu v reálném světě, řídí soustavy akcí a komplexního chování při interakci s prostředím a touto interakcí jsou schémata zpětně determinována. Fenotypní schémata představují momentální projevy chování a jsou ztělesněny akcemi organismu v prostředí. Spouštěčem genotypních schémat je bottom-up proces, který vzniká na situačním základě, fenotypní schémata jsou potom založena na zkušenostech a očekávání jedince, kde se už jedná o top-down procesy (Plant & Stanton, 2012). Hlavní kognitivní struktury, které ovlivňují tento proces, se nazývají anticipační schémata (Chimir et al., 2005).

Jak uvádí Cacciabue (2004), pouze málo výzkumného zájmu je věnováno dynamické povaze interakce člověk-počítač, zatímco statická perspektiva je dobře explorována a zahrnuta ve většině přístupů. Moderní technologie nabízí možnost simulace prostředí, kde je zároveň možné přesné měření interakce člověka a takového prostředí a otevírají se nové možnosti pro výzkum. Interakce s VR je vystavěna primárně na vizuálně percepčním základě. Neiserův percepční cyklus (Neisser, 1967) zdůrazňuje senzomotorickou povahu percepce a kognice právě s ohledem na anticipační schémata. Anticipační schémata představují dynamické struktury, které ovlivňují proces vyhledávání informací v prostředí, ve kterém se člověk pohybuje. Cyklicky, s ohledem na přichodící informaci z prostředí, motorická aktivita selektivně zaměřuje pozornost na specifické nápovědi ve vizuálním poli. Ačkoliv předchozí studie srovnávající rozdílné typy 3D vizualizace mnohdy neprokázaly percepční rozdíly jako signifikantní (Fabrikant et al., 2014; Popelka & Brychtová, 2013; Wood et al., 2005), náš dosavadní výzkum vizualizací prezentovaných ve virtuální realitě poukazuje na odlišnosti při samotné interakci s virtuálními expozicemi (Špriňarová a kol., 2015). Rozdílné typy 3D zobrazení, které nabízí odlišné vizuální nápovědi (Buchroithner & Kunst, 2013) se liší na úrovni kognitivní elaborace. Mohou tedy vyvolat odlišný typ motorické akce a v závislosti na tom také odlišnou percepci u informačně ekvivalentních obsahů. Jak naznačuje Špriňarová a kol. (2015), lze předpokládat, že způsob prezentace informace ve 3D virtuálním prostředí a další faktory, jakými jsou imerze do VR a typ ovládání interaktivní expozice, které do interakce člověka s 3D prostředím vstupují, ovlivňují mentální zpracování, pochopení, zapamatování a vyhodnocení prezentované informace. Lidské pochybení, které představuje zásadní téma v mnoha aplikovaných oblastech průmyslu, úzce souvisí s percepčním cyklem. Pochybení člověka-operátora je způsobeno právě opomenutím určitého aspektu dané situace, což souvisí jak s kognitivní, tak se senzomotorickou stránkou lidské psychiky. Při designu 3D rozhraní je tedy na místě zvážit, zdali je zvyšující se pravděpodobnost lidského pochybení za cenu například úspory času, statisticky lepšího výsledku práce, popřípadě většího vnošení do VR, objektivně na místě. Náš dosavadní výzkum zaměřený na schopnost člověka identifikovat prostorové parametry virtuálních 3D geografických modelů naznačil tendenci uživatelů v Reálném 3D zobrazení opomínat ve větší míře některé důležité aspekty scény (Juřík et al., 2016). Tato tendence pro nás představuje novou výzkumnou linii 3D zobrazení, ve které se zaměřujeme na psychologické aspekty interakce člověk počítač optikou lidského faktoru. Vzhledem k výše diskutované teorii percepčního cyklu a proximálních nápovědí se snažíme postihnout senzomotorickou aktivitu účastníků při interakci s virtuálním prostředím. V případě, že se konkrétní rozhraní ukáže sice jako rychlejší a přesnější,

nicméně se zvýšenou tendencí k lidskému pochybení, je nezbytné tento fakt zvážit při aplikaci daného designu rozhraní, které bude např. sloužit člověku při řešení život ohrožujících událostí. Pro výzkumné účely této problematiky navrhujeme konkrétní výzkumné rozhraní.

4 Jak měřit lidskou chybu

Interakce s UI je dynamický proces. Pro odhalení konsekvencí, jaké s sebou přináší virtuální interaktivní vizualizace, je nezbytné stanovit parametry výzkumného prostředí tak, aby byl vyvolán žádoucí typ interakce a zároveň zachyceny sledované parametry s přihlédnutím k výše zmíněnému percepčnímu cyklu. Pro výzkumné účely je v rámci VR možné využít široké škály moderních technologií, které dokáží zajistit přesné měření uživatelského chování, např. typ a kvantitu motorické aktivity uživatele, fokus zájmu ve vizuálním poli nebo kvantitu pohybu ve virtuálním prostoru. Na jedné straně je nezbytné zajistit formy výstupů z VR – primárně tedy formy vizualizace. Na straně druhé potom stojí cesty, kterými člověk do interakce vstupuje – tedy jak ovládá a manipuluje s virtuálním prostředím. Na základě našeho výzkumu navrhujeme interface pro výzkum lidské chyby ve 3D rozhraní, který umožňuje optimální interakci s virtuálním rozhraním a zároveň umožňuje sběr dat pro následnou analýzu.

Výstupy Pro prezentaci podnětového materiálu je vhodné využít širokoúhlé 3D projekční plátno, které zajišťuje stereoskopii. 3D vizualizace je potom tvořena grafickým provedením, které simuluje monokulární nápovědi hloubky v expozici. Takto je zajištěno Pseudo 3D zobrazení. V případě Reálného 3D je nezbytné zajistit stereoskopii za využití periferního zařízení. Jako optimální se jeví využití aktivních Shutter 3D brýlí, které emulují přesvědčivý 3D vjem a zároveň jsou jednoduché a praktické (Torres et al., 2013). Na další úrovni je možné využít Head-mounted displeje, což jsou zařízení pro emulaci imerzivní VR. Sledování oblastí zájmu účastníků ve vizuálním poli (tzv. areas of interest) je u obou zmíněných technologií možné monitorovat za použití eye-tracking systému.

Vstupy Pro samotnou interakci s prostředím je třeba zajistit zařízení, které nabízí vedle možnosti sledování lidské motorické aktivity také efektivní možnosti ovládání virtuální expozice. Zároveň by toto zařízení člověka nemělo významně limitovat, mělo by ideálně zajišťovat 6 DoF (6 stupňů volnosti pohybu). Pro tento účel je možné sestavit jako kontrolní a měřicí zařízení kombinaci Motion Capture systému a Wii RC ovladače. Wii RC slouží jako akční tlačítko a samotný pohyb a manipulaci s prostředím zajišťuje přesný MoCap systém. Pro následné měření je nutné vydefinovat, zdali bude cílem měření samotný pohyb účastníka před projekčním plátnem, nebo zdali je má

být výzkumným záměrem měřit jeho pohyb ve virtuálním prostředí – např. pohyb virtuální kamery, se kterou operátor manipuluje.

Podnětový materiál Jako vhodný podnětový materiál se jeví geografická reprezentace prostorových dat, protože mapa obsahuje přesně definované a dobře měřitelné množství informace, kterou je člověk schopen extrahovat — mapa představuje pro psychologický výzkum ideální výzkumný stimulus díky přesně kontrolovatelnému počtu vizuálních návodů.

Pro emulaci imerzivních virtuálních rozhraní je možné využít software, jakými jsou VRECKO a Unity 3D software. VRECKO je open-source software, který je vyvíjen od roku 2003 HCI Laboratoří na Fakultě informatiky, MU. VRECKO je naprogramováno v C++ a podporuje různé zařízení pro interakci člověk-počítač (3D myš, Wii RC). Je možné ho propojit s Motion Capture systémem. VRECKO nabízí také různé možnosti 3D projekce (stereoskopické/pseudo 3D projekce), pro více detailů <http://vrecko.cz>. Unity 3D engine je platforma vyvíjená Unity Technologies, která podporuje zapojení Head-mounted displejů (např. Oculus Rift). Tyto platformy je možné individuálně přizpůsobit pro sběr a následnou analýzu dat.

Cílem výzkumu interakce člověk-počítač v návaznosti na senzomotorický cyklus je zachytit souvislost mezi tím, jaký typ návodů přitahuje pozornost operátora a typem akce, kterou operátor na základě těchto návodů vykoná. Kombinace zmíněných zařízení a stimulů zajišťuje možnost studovat interakci člověka s rozhraním také právě s ohledem na Neisserovu teorii. Aspirací je kontrola anticipačních schémat v rámci interakce člověk-počítač. Pomocí eye-trackeru lze v rámci rozhraní definovat oblasti zájmu operátora na 3D plátně, popřípadě také ve vizuální expozici prezentované prostřednictvím Head-mounted displeje. V návaznosti na oblast zájmu lze usoudit, které návodů vedly operátora k dané akci a jaké další návodů si svojí akcí zpřístupnil. Sběr přesných dat je zajištěn ze strany software. Toto rozhraní tedy splňuje parametry pro měření senzomotorické interakce s VR, jak je diskutována v oblastech lidského faktoru (Plant & Stanton, 2012).

5 Diskuze a závěr

Článek shrnuje některé poznatky o výzkumu interakce člověka s virtuálním rozhraním a navrhuje optimální parametry VR rozhraní pro psychologický výzkum této problematiky. V případě měření interakce člověk-počítač se jedná o dvousměrný proces vstupů a výstupů, které se navzájem ovlivňují v neustálém cyklu. V případě výzkumu lidské chyby, která je v mnoha oblastech zásadní oblastí výzkumného zájmu, je na základě argumentů uvedených výše nezbytná

analýza procesů a okolností, které k selhání lidského faktoru vedly. Výzkumníci se shodují na tom, že neexistuje žádná jednoduchá příčina lidského selhání (Plant & Stanton, 2012; Dekker, 2006; Hall & Silva, 2008; Reason, 1990), a že pro analýzu příčin nehod a incidentů je nezbytné studovat senzomotorický proces interakce člověka s prostředím.

Pro tento účel navrhuje zohlednit Neisserův model percepčního cyklu (1967) s důrazem na anticipační schémata, coby determinanty senzomotorické interakce člověk-počítač. Abychom mohli postihnout a měřit zahrnuté procesy, je nezbytné využít dostupné technologie, které dokáží na jedné straně věrně simulovat zkoumané prostředí, a zároveň umožňují přesné měření behaviorální odezvy člověka-operátora pracujícího v tomto rozhraní. Optimální rozhraní je možné vytvořit kombinací technologií jako eye-tracker, Motion Capture systém, 3D projekční plátno a patřičná softwareová platforma.

Poděkování

Příspěvek vznikl v rámci projektu MUNI/M/0846/2015 “Influence of cartographic visualization methods on the success of solving practical and educational spatial tasks”.

Literatura

Amalberti, R., (2001). The paradoxes of almost totally safe transportation systems. *Safety Science* 37, 109–126.

Anderson, R. C., (1977). The notion of schemata and the educational enterprise. In: Anderson, R.C., Sapiro, R.J., Montague, W.E. (Eds.), *Schooling and the Acquisition of Knowledge*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 415–431.

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators & Virtual Environments*, 15(4), 359e372.

Bennett, S. (2001). *Human Error – By Design?* Perpetuity Press, Leicester.

Björk, S., Holopainen, J. (2004). *Patterns In Game Design*. Charles River Media.

Boughzala, I., de Vreede, G. J., Limayem, M. (2012). Team Collaboration in Virtual Worlds: Editorial to the Special Issue. *Journal of the Association for Information Systems*. 13 (Special Issue): 714-734.

- Bryson, S. (1999). Virtual Reality: A Definition History, Omnibus Lexicon Definition Supplement.
- Buchroithner, M. F., Knust, C. (2013). True-3D in Cartography—Current Hard and Softcopy Developments. In Moore A, Drecki I: Geospatial Visualisation, Berlin, Heidelberg: Springer, 2013: 41 - 65.
- Cacciabue, P. (2004). Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training. *Reliability Engineering & System Safety*, 83(2), 229-240.
- Chalmers, P. A. (2003). The role of cognitive theory in human-computer interface. *Computers in Human Behaviour* 19, 593–607.
- Chimir, I. A., Abu-Dawwas, A. A., Horney, M. A. (2005). Neisser's cycle of perception: formal representation and practical implementation. *Journal of Computer Science* 10, 106–111 (special issue).
- Davis, A., Murphy, J., Owens, D., Khazanchi, D., Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems*, 10(2): 90-117.
- Dekker, S. (2002). The Re-Invention of Human Error. Technical Report 2002-01. Lund University School of Aviation.
- Dekker, S. (2006). The Field Guide to Understanding Human Error. Aldershot, Ashgate.
- Endsley, M., R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors* 37: 32-64.
- Fabrikant, S. I., Maggi, S., & Montello, D. R. (2014). 3D Network spatialization: does it add depth to 2D representations of semantic proximity?. In *Geographic Information Science* (pp. 34-47). Springer International Publishing.
- Fedota, J., Parasuraman, R., 2009. Neuroergonomics and human error. *Theoretical Issues in Ergonomics Science* (iFirst), 1–20.
- Forsell, C. (2007): Perceptually Motivated Constraints on 3D Visualizations, PhD dissertation, Uppsala University.
- Goel, L., & Prokopec, S. (2009). The socialness of virtual worlds. *AMCIS 2009 Proceedings*.
- Hall, J. G., Silva, A. (2008). A conceptual model for the analysis of mishaps in human-operated safety-critical systems. *Safety Science* 46, 22–37.
- Hammond, K. and Stewart, T. (2001). *The essential Brunswick*. Oxford: Oxford University Press.
- Hankiewicz, K. (2012). Ergonomic characteristics of software for enterprise management systems. In: P.Vink (ed.) *Advances in Social and Organizational Factors*, CRS Press, Boca Ranton, pp 279-287.
- Hammond, K. Stewart, T. (2001). *The essential Brunswick*. Oxford: Oxford University Press.
- Hassell, M., & Limayem, M. (2010). Working in the new way: A preliminary study of media synchronicity and job satisfaction. *AMCIS 2010 Proceedings*.
- Hirmas, D. R., Slocum, T., Halfen, A. F, White, T., Zautner, E., Atchley, P., Liu, H., Johnson, W. C., Egbert, S., McDermott, D. (2014). Effects of Seating Location and Stereoscopic Display on Learning Outcomes in an Introductory Physical Geography Class. *Journal of Geoscience Education*. 62(1): 126-137.
- Hochmitz, I., Yuviler-Gavish, N. (2011). Physical Fidelity Versus Cognitive Fidelity Training in Procedural Skills Acquisition. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5): 489-501.
- Hollywell, P. D. (1996). Incorporating human dependent failures in risk assessments to improve estimates of actual risk. *Safety Science* 22, 177–194.
- Juřík, V., Herman, L., Šašinka, Č., Stachoň, Z., Chmelík, J. (2016). Is more less? The Psychological Perspective on Three-dimensional interaction with virtual geographical environments. *Nepublikovaný manuskript*.
- Kontogiannis, T., Malakis, S. (2009). A proactive approach to human error detection and identification in aviation and air traffic control. *Safety Science* 47, 693–706.
- Lange, M., Hjalmarsson, J., Cooper, M., Ynnerman, A., Duong, V. (2003). 3D Visualization and 3D Voice Interaction in Air Traffic Management. *SIGGRAD*.
- Larkin, J. H., Simon, H., A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, 1987, 11(1): 65-100. ISSN 1551-6709.
- Lipshitz, R., Klein, G., Orasanu, J., Sala, E. (2001). Focus article: taking stock of naturalistic decision making. *Journal of Behaviour Decision Making* 14, 331–352.

Finančníci - pragmatici?

Konverzačný kontext a doménová špecifickosť v posudzovaní rizika a rozhodovaní

Lenka Kostovičová, Katarína Dudeková

Ústav experimentálnej psychológie CSPV SAV
Dúbravská cesta 9, 84101 Bratislava
lenka.kostovicova@savba.sk, katarina.dudekova@savba.sk

Abstrakt

Interpretácia účelu použitia slovných vyjadrení pravdepodobnosti ovplyvňuje vnímanie rizika. V rámci kontrolnej skupiny (neutrálny zámer) alebo jednej z dvoch experimentálnych podmienok (zdvorilostné zábery) účastníci, finanční profesionáli, posudzovali pravdepodobnosť, negatívu a intenciu ku konaniu v dvoch doménach. Konverzačný kontext neovplyvnil subjektívnu úroveň rizika, závažnosti a tendencie k správaniu. Identický verbálny kvantifikátor však viedol k rozdielnemu posúdeniu pravdepodobnosti a odlišným voľbám, v závislosti od domény. Vo finančnej doméne rod moderoval vplyv vnímanej negativity na tendenciu predať akcie, pričom sa efekt prejavil len u žien. Venujeme sa limitom generalizácie predošlých zistení a praktickým aspektom komunikácie informácií o riziku.

1 Úvodná kapitola

Predstavte si, že Vám lekár povie: "Pravdepodobne budeme musieť v budúcnosti podstúpiť operáciu". Ako by ste si vysvetlili tento vážny odhad? Mysleli by ste si, že lekár vyjadruje svoju subjektívnu mieru neistoty? Prípadne, že Vás chce ušetriť pred krutou realitou a prezentuje informáciu taktne a ohľaduplne? Alebo by ste mali tendenciu veriť, že chráni predovšetkým samého seba pre prípad, že sa jeho prognóza nenaplní a mohol by niesť vinu za chybný odhad? Aký zámer hovorcovi prisúdime ovplyvní mieru pravdepodobnosti a závažnosti naplnenia sa odhadu a následne môže viesť k zmenám v našom správaní.

1.1 Verbálne pravdepodobnostné kvantifikátory

Ľudia chápu slovné vyjadrenia, ako "je možné", či "je skoro vylúčené", rôzne. Pre jedného človeka môže znamenať pojem "pravdepodobne" hodnotu okolo 60%, pre iného 80%. Dôležitú rolu zohráva i interpretácia účelu použitia slovného výrazu, ako sme opísali v úvodnom príklade. Je zrejmé, že pri používaní slovných odhadov sa vytvára veľký priestor pre nedorozumenia (Karelitz a Budescu, 2004).

Používanie verbálneho pravdepodobnostného odhadu by sa preto mohlo zdať nevhodné. Avšak, vo veľkej časti prípadov nie je presný numerický údaj možné poskytnúť. Navyše, Sirota a Juanchich (2012) upozorňujú, že číselné odhady taktiež podliehajú efektu interpretácie zámeru komunikátora. Pozrime sa bližšie na rôzne možnosti pochopenia účelu použitia slovných pravdepodobnostných kvantifikátorov.

1.2 Interpretácia zámeru použitia verbálnych pravdepodobnostných kvantifikátorov

Verbálne výrazy môžu slúžiť na vyjadrenie subjektívneho odhadu v súvislosti s potenciálnym výsledkom udalosti. Človek komunikuje mieru neistoty, akú reálne vníma. Zámer nazývame nástrojom komunikácie pravdepodobnosti (Juanchich a spol., 2012). Použitím slovného kvantifikátora však môžeme sledovať i pragmatické záujmy.

Pragmatické účely súvisia s ochranou sociálnej tváre. Sociálna tvár predstavuje obraz a predstavy o človeku - jeho "self-image". Uplatňuje sa pri sociálnej interakcii, jednotlivec ju zastáva pomocou neverbálnych a verbálnych aktov, pričom sa očakáva súlad medzi jeho správaním a sociálnym obrazom. Sociálny obraz je pozitívna sociálna hodnota, preto sa ju ľudia snažia chrániť. Rovnako však pri spoločenskom kontakte ochraňujú aj sociálnu tvár iných. Dôvodom je emocionálne spojenie so sociálnym obrazom ostatných, morálne dôvody alebo snaha vyhnúť sa nepríjemnostiam (Goffman, 2003).

Nástroj ochrany tváre príjemcu sa uplatňuje, ak chce hovorca chrániť emócie, respektíve tvár ostatných (t.j. príjemcov; Bonnefon a Villejoubert, 2005; vid' taktiež Juanchich a spol., 2012, Sirota a Juanchich, 2012). Z toho pramení snaha o taktnú, citlivú, komunikáciu pravdepodobnosti. Nástroj ochrany tváre hovorcovi sa týka situácií, kedy komunikátor preventívne zmierňuje svoj odhad, aby neniesol vinu v prípade jeho nenaplnenia, čiže chráni svoju vlastnú tvár (Juanchich a spol., 2012). Keďže pri presnom numerickom odhade

je riziko nenaplnenia vyššie, je na mieste očakávať výskyt práve tohto účelu.

1.3 Vplyv interpretácie zámeru použitia verbálnych pravdepodobnostných kvantifikátorov na posúdenie rizika

Ak ide o negatívnu prognózu, interpretácia použitia slovného výrazu v zmysle pragmatického účelu vedie k zvýšeniu úrovne vnímanej pravdepodobnosti naplnenia výroku v porovnaní s neutrálnym zámerom. Ak si prijímatelia myslia, že komunikátor chce vyjadrením chrániť seba alebo byť voči nim taktný, majú tendenciu veriť, že hovorca vyjadruje nižšiu mieru rizika, než má reálne na mysli (Juanchich a spol., 2012).

Skúmaním situácií použitia slovného vyjadrenia pravdepodobnosti s úmyslom chrániť príjemcu a v neutrálnom kontexte Bonnefon a Villejoubert (2005) zistili, že prvý typ situácií viedol k vyššej miere vnímaného rizika naplnenia odhadov. Juanchich a spol. (2012) rozšírili možnosti interpretácie o nástroj ochrany hovorcu. Zistili, že ide o najčastejší spôsob vnímaného zámeru. Účel ochrany tváre hovorcu viedol v ich experimentoch u negatívnych prognóz k najvyššej miere vnímanej pravdepodobnosti. Navyše, vyššia úroveň vnímaného rizika viedla k väčšej ochote zmeniť správanie, konkrétne predat' akcie.

1.4 Expertnosť v posudzovaní a rozhodovaní

V posudzovaní pravdepodobnosti častokrát zlyhávajú i experti (Gigerenzer a Hoffrage, 1995; Goodman-Delahunty a spol., 2010). Profesionáli v oblasti financií sú vo svojich odhadoch miscalibrovaní a nadmerne optimistickí (Glaser a kol., 2005; Grežo, 2015). Avšak výsledky nášho nedávneho výskumu naznačujú, že expertnosť by mohla znižovať efekt kognitívnych odchýlok v usudzovaní (Dudeková & Kostovičová, 2015). Preto sme sa rozhodli uskutočniť experiment so vzorkou finančných profesionálov a overiť, nakoľko sú imúnni či náchylní podliehať pri posudzovaní rizika vplyvu interpretácie zámeru hovorcu.

1.5 Cieľ výskumu

V rámci upravenej repliky štúdie Juanchich a kol. (2012) je našim cieľom zistiť, či má interpretácia účelu slovného vyjadrenia pravdepodobnosti vplyv na vnímanie rizika a negativity a na intenciu ku konaniu. Tieto efekty nás zaujímajú vzhľadom k dvom doménam, z ktorých jedna predstavuje oblasť expertízy participantov. Predpokladáme, že efekt účelu použitia slovného odhadu bude významný len v oblasti zdravia a nie v oblasti financií. Pozrieme sa i na ďalšie potenciálne prediktory rozhodovania o zdraví a financiách: dĺžka praxe, rod, vek a poradie administrovaných materiálov.

2 Metódy

2.1 Participanti

Nášho výskumu sa zúčastnilo 99 finančných profesionálov vo veku 21 až 62 rokov ($M = 37,5$; $SD = 9,4$). Podiel mužov (54,5%) a žien (45,5%) bol vyrovnaný. Dĺžka praxe sa pohybovala v rozmedzí 0,5 až 34 rokov ($M = 9,3$; $SD = 8,1$). Medzi profesiami participantov sa najčastejšie objavili nasledovné: finančný konzultant, finančný poradca, finančný sprostredkovateľ, finančný agent, finančný analytik, účtovník a ekonóm.

2.2 Dizajn a operacionalizácia

Participanti boli náhodne priradení do jednej z troch skupín: kontrolnej s neutrálnym zámerom ($n = 36$) alebo jednej z experimentálnych podmienok s pragmatickým zámerom ($n = 35$, $n = 28$).

2.3 Materiály a procedúra

Najskôr si participanti prečítali scenár:

i) **Financie:** *Investovali ste značnú časť svojich úspor do burzy cenných papierov v presvedčení, že Vaše akcie budú výnosné. Následne sa o investícii bavíte s Vaším známym, ktorý je burzový maklér.*

ii) **Zdravie:** *Ste na pravidelnej preventívnej zdravotnej prehliadke, kde zistíte, že máte zvýšenú hladinu cholesterolu v krvi. O tomto výsledku sa následne bavíte s Vaším lekárom.*

Následne im bola navodená interpretácia zámeru hovorcu:

i) **Nástroj komunikácie pravdepodobnosti:** *Viete o ňom, že sa spravidla vyjadruje vecne a jasne.*

ii) **Nástroj ochrany tváre hovorcu:** *Viete o ňom, že sa spravidla vyjadruje opatrne, aby neniesol vinu v prípade, že sa jeho názor ukáže byť chybný.*

iii) **Nástroj ochrany tváre príjemcu:** *Viete o ňom, že sa spravidla vyjadruje taktne, aby človeka nezranil alebo neurazil.*

Ďalej si prečítali výrok obsahujúci verbálny pravdepodobnostný kvantifikátor:

i) **Financie:** *Je možné, že hodnota Tvojich akcií klesne.*

ii) **Zdravie:** *Je možné, že Vám v budúcnosti bude hroziť nebezpečenstvo srdcovo-cievnej príhody.*

Úlohou participantov bolo vyjadriť vnímanú úroveň pravdepodobnosti a negativity informácie, ako aj mieru tendencie k správaniu:

i) **Subjektívna pravdepodobnosť:** *Na základe tejto informácie, akú pravdepodobnosť pripisujete tomu, že hodnota Vašich akcií klesne? / ..., že Vám v budúcnosti bude hroziť nebezpečenstvo srdcovo-cievnej príhody?*

ii) Subjektívna negativita: *Ako vnímate informáciu, ktorú ste sa dozvedeli od Vášho známeho? / ... od Vášho lekára?*

iii) Tendencia k správaniu: *Čo by ste, po vypočítaní informácie od Vášho známeho, urobili s Vašimi akciami? / ... s Vašimi stravovacími návykmi?*

Pravdepodobnosť hodnotili finančníci na škále 0% až 100%. Negativitu posudzovali na Likertovej škále, kde 1: *maximálne negatívne* a 10: *maximálne pozitívne* (odpovede sme rekódovali, aby vyššia hodnota znamenala vyššiu negativitu). Tendencia ku konaniu mala v doméne financií koncové body: 1: *určite by som si ich [akcie] ponechal/a* a 10: *určite by som ich [akcie] predal/a*. U zdravia išlo o hodnoty 1: *určite by som ich [stravovacie návyky] nezmenil/a* a 10: *určite by som ich [stravovacie návyky] zmenil/a*.

Poradie administrácie scenárov o financiách a zdraví bolo náhodné. Po vyplnení všetkých, vrátane socio-demografických, údajov, mali účastníci možnosť v osobitnom dokumente uviesť svoju emailovú adresu, kde im bola neskôr doručená správa z výskumu.

3 Výsledky

Subjektívna pravdepodobnosť, negativita a tendencia k správaniu nespĺňali podmienky normality, preto sme pracovali s mediánmi, interkvartilovými rozpätiami a neparametrickými testami. Poradie, dĺžka praxe a vek nehrali rolu v rámci hodnôt závislých premenných.

3.1 Efekt interpretácie

Hodnoty pravdepodobnosti, závažnosti a intencie konať uvádzame v Tabuľkách 1 (financie) a 2 (zdravie). Riadky predstavujú jednotlivé zámery. NKP znamená nástroj komunikácie pravdepodobnosti, NOH nástroj ochrany hovorcu a NOP nástroj ochrany príjemcu.

| | Pravdepodobnosť | | Negativita | | Správanie | |
|-----|-----------------|------|------------|-----|-----------|-----|
| | Mdn | IQR | Mdn | IQR | Mdn | IQR |
| NKP | 50,0 | 40,0 | 6,0 | 4,5 | 3,5 | 4,8 |
| NOH | 60,0 | 41,7 | 7,0 | 2,0 | 5,0 | 3,0 |
| NOP | 67,5 | 23,8 | 6,5 | 1,8 | 5,0 | 4,0 |

Tab. 1: Výsledky v doméne financií

| | Pravdepodobnosť | | Negativita | | Správanie | |
|-----|-----------------|------|------------|-----|-----------|-----|
| | Mdn | IQR | Mdn | IQR | Mdn | IQR |
| NKP | 50,0 | 22,5 | 6,5 | 2,8 | 7,5 | 5,0 |
| NOH | 60,0 | 30,0 | 7,0 | 2,0 | 8,0 | 4,0 |
| NOP | 60,0 | 40,0 | 7,0 | 2,0 | 8,0 | 2,8 |

Tab. 2: Výsledky v doméne zdravia

Rozdiely vzhľadom na zámer, overované Kruskal-Wallisovým testom, neboli významné ani u jednej

zo šiestich premenných. Párové porovnania Mann-Whitneyho U-testmi, aplikujúc Bonferroniho korekciu, taktiež neprinesli významné zistenia. Interpretácia zámeru komunikátora nemala v našom výskumnom súbore vplyv na posúdenie pravdepodobnosti a negativity, ani na vyjadrenú tendenciu k správaniu.

3.2 Porovnanie domén

Identický výraz - "je možné" - viedol u dvoch domén k rozdielnej miere vnímaného rizika naplnenia sa scenára a k odlišnej úrovni tendencie k správaniu. Finančný scenár ($Mdn = 55,0$; $IQR = 35,0$) sa účastníkom zdal viac pravdepodobný než zdravotný ($Mdn = 50,0$; $IQR = 40,0$), $Z = -2,5$; $p = 0,012$; $rm = 0,25$. Po prečítaní výroku lekára ($Mdn = 8,0$; $IQR = 4,0$) však mali výrazne vyššiu tendenciu zmeniť správanie ako po prečítaní výroku od burzového makléra ($Mdn = 5,0$; $IQR = 5,0$), $Z = -7,1$; $p < 0,001$; $rm = 0,58$.

3.3 Ďalšie faktory

Pozreli sme sa i na interakcie faktorov, vrátane kontrolných premenných. Zistili sme rodové rozdiely v oblasti financií. Ženy vnímali subjektívnu pravdepodobnosť a negativitu odhadu o strate hodnoty akcií vyššie ako muži a mali i vyššiu tendenciu akcie predat'. Detailné výsledky poskytujeme v Tabuľke 3.

| | Pravdepodobnosť | | Negativita | | Správanie | |
|------|-----------------|------|-------------|-----|-------------|-----|
| | Mdn | IQR | Mdn | IQR | Mdn | IQR |
| muži | 50,0 | 41,3 | 6,0 | 2,3 | 3,5 | 5,0 |
| ženy | 70,0 | 30,0 | 7,0 | 2,0 | 5,0 | 4,0 |
| test | $U = 851,0$ | | $U = 663,0$ | | $U = 880,5$ | |
| | $p = 0,010$ | | $p < 0,001$ | | $p = 0,017$ | |
| | $rm = 0,26$ | | $rm = 0,40$ | | $rm = 0,24$ | |

Tab. 3: Rodové rozdiely v doméne financií

V rámci moderačnej analýzy sme zistili, že posúdená negativita informácie od burzového makléra viedla k tendencii predat' akcie leu u žien (95% CI [0,10; 0,98]) a nie u mužov (95% CI [-0,12; 0,52]). Prediktory intencie k správaniu v oblasti zdravia sa nám identifikovať nepodarilo.

4 Diskusia

Posudzovanie rizika a pravdepodobnosti je dôležitou súčasťou našich každodenných rozhodnutí, avšak i volieb so závažnými dôsledkami – napríklad v doméne zdravia a financií. Jedným zo zdrojov skreslení vo vnímaní rizík sú domnienky o zámere hovorcu v prípade, že ten komunikuje odhad pravdepodobnosti použitím slovného kvantifikátora, napr. "je možné, že na pohovore uspeješ". Otázkou je, nakoľko je podliehanie efektu prisúdeného účelu použitia slovných

odhadov doménovo špecifické a podmienené expertnosťou.

Vychádzajúc z nedávnych výskumov o verbálnych pravdepodobnostných kvantifikátoroch (Bonneton a Villejoubert, 2005; Juanchich a spol., 2012; Karelitz a Budescu, 2004; Sirota a Juanchich, 2012) sme v našej štúdiu experimentálne testovali vplyv interpretácie zámeru na posúdenie rizika, závažnosti a tendencie ku konaniu. Finanční profesionáli sa vyjadrovali k dvom témam - zdravotnej (odhad rizika srdcovo-cievnej príhody) a finančnej (odhad rizika straty akcií).

Na rozdiel od východiskových štúdií, kde pragmatický účel použitia slovných odhadov viedol k vnímaniu vyššej úrovne pravdepodobnosti, negativity a tendencie k správaniu, nám sa efekt interpretácie zámeru hovorca nepodarilo potvrdiť. Zdá sa, že v rámci úsudkov a volieb finančných profesionálov nezohráva zámer hovorca kľúčovú úlohu. Finančníci teda podliehajú niektorým kognitívnym skresleniam do menšej miery než laici (napr. Dudeková & Kostovičová, 2015).

Zároveň sme zistili, že posúdenie verbálnych pravdepodobností je doménovo špecifické, pretože rovnaký výraz viedol k vnímaniu vyššieho rizika naplnenia odhadu o zmene hodnoty akcií a zároveň k vyššej ochote zmeniť svoje stravovacie návyky. Je možné, že finančníci reagujú na zmenu hodnôt akcií menej citlivo, keďže v rámci skúseností vo svojej profesii vedia, že to nie je výnimočný jav. Preto nepreferujú okamžitý predaj akcií. Avšak alarmujúca informácia o hrozbe srdcovo-cievnej príhody ich vedie k väčšej ochote zväziť zmenu stravovania, hoci si pod ňou môžu predstaviť len miernu úpravu jedálneho lístka.

Zdá sa, že ani finanční profesionáli nie sú homogénna skupina - prejavili sa u nich totiž výrazné rodové rozdiely. Ženy vnímali naplnenie odhadu o strate hodnoty investície ako viac pravdepodobné a závažné a v dôsledku vnímanej negativity prognózy by na rozdiel od mužov spravidla pristúpili k predaju akcií. Výsledok je v súlade s evidenciou o rodových rozdieloch v postoji k riziku (pre prehľad pozri Dudeková, 2014).

Aj náš výskum potvrdzuje, že ľudia chápu verbálne vyjadrenia pravdepodobnosti rôzne. Aby sme sa dokázali rozhodovať efektívne, je pred vykonaním voľby podporená určitým odporúčaním dobrá mať na pamäti princípy zdvorilosti, ktoré môžu ovplyvniť tak vyjadrenie hovorca, ako aj porozumenie príjemcu informácie. Zistenia výskumov o posudzovaní rizika a pravdepodobnosti je možné využiť v rámci komunikácie informácií cieľovým skupinám, pri tvorbe verejných politík a intervenciách v súvislosti s optimalizáciou spoločensky relevantných rozhodnutí.

Pod'akovanie

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

- Bonneton, J. F. a Villejoubert, G. (2005). Communicating likelihood and managing face: Can we say it is probable when we know it to be certain? V *Proceedings of the 27th annual conference of the cognitive science society*, str. 316–321.
- Dudeková, K. (2014). *Psychologické aspekty parametrov rozhodovania v prospektovej teórii*. Bratislava: Univerzita Komenského.
- Dudeková, K. a Kostovičová, L. (2015). Oscary, HDP a CRT: Efekt ukotvenia u finančných profesionálov v kontexte doménovej špecifickosti a kognitívnej reflexie. V zborníku *Kognícia a umelý život XV*, str. 50–56.
- Gigerenzer, G. a Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*. 102(4): 684-704
- Glaser, M., Langer, T. a Weber, M. (2005). *Overconfidence of professionals and lay men: Individual differences within and between tasks?* University of Mannheim.
- Goffman, E. (2003). On face-work: An analysis of ritual elements in social interaction. *Reflections*. 4(3): 7–13
- Goodman-Delahunty, J., Granhag, P. A., Hartwig, M. a Loftus, E. F. (2010). Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy, and Law*. 16(2): 133–157
- Grežo, M. (2015). Efekt nadmernej sebadôvery v usudzovaní a rozhodovaní finančných profesionálov. V zborníku *Cesty k rozhodovaniu: Šesť dekád psychologického výskumu*, str. 66–75.
- Juanchich, M., Sirota, M. a Butler, C. L. (2012). The perceived functions of linguistic risk quantifiers and their effect on risk, negativity perception and decision making. *Organizational Behavior And Human Decision Processes*. 118(1): 72–81
- Karelitz, T. M. a Budescu, D. V. (2004). You say "probable" and I say "likely": Improving interpersonal communication with verbal Probability phrases. *Journal Of Experimental Psychology Applied*. 10(1): 25–41

Sirota, M. a Juanchich, M. (2012). To what extent do politeness expectations shape risk perception? Even numerical probabilities are under their spell! *Acta Psychologica*. 141(3): 391–399

Who's moving it? Sense of agency in schizophrenia

Eva Kozáková, Eduard Bakštejn, Filip Španiel, Tereza Nekovářová, Jan Rydlo

Národní ústav duševního zdraví

Topolová 748, 250 67 Klecany

eva.kozakova@nudz.cz, eduard.bakstein@nuzd.cz, filip.spaniel@nudz.cz,

tereza.nekovarova@nudz.cz, jan.rydlo@nudz.cz

Petra Šustová

Katedra psychologie, Filozofická fakulta UK

Nám. J. Palacha 2, 116 38 Praha 1

Abstract

Sense of Agency (SoA) is a term depicting the ability to distinguish between self- and other-produced actions. In schizophrenia, first-rank symptoms - such as delusion of control or auditory hallucinations - reflect disturbances in SoA. In our research we studied SoA in first-episode schizophrenia patients (FES) and healthy controls (HC) in an original Ghost city task (GCT), which tests the ability to correctly distinguish between movements that are fully under one's control from those that are influenced. We compared performance of both groups, and in patients we tested the relationship between the performance in GCT and scores in PANSS or presence of the Schneider first-rank symptoms.

1 Introduction

Phenomenological analyses of schizophrenia symptoms indicate that disturbances of Self might be a core marker of the disease (Sass, 2014; Sass & Parnas, 2003). In the 1960s, Kurt Schneider described the disturbance of ego boundaries as a hallmark of schizophrenia and used the term first-rank symptoms to account for experiences such as auditory hallucinations (e.g. commenting voices), abnormal perception which blurs the barrier separating self from the environment (e.g. thought insertion, delusion of control), and delusional perception (Strauss & Carpenter, 2013).

It has been shown that alterations in SoA - the capacity to correctly distinguish self-produced from other-produced actions - that are pronounced in schizophrenia symptoms are detectable also in patients' performance in behavioural tests. This alteration is reflected in diminished capacity of self-monitoring of movements (Frith, 2012; Synofzik, Thier, Leube, Schlotterbeck, & Lindner, 2010; Voss et al., 2010). According to a well-established forward model of control, the reason for the alteration in SoA is a mismatch between predictions of a motor action and actual sensory consequences (Blakemore, Wolpert, & Frith, 2002; Frith, 1992). A more general and multilevel account of perceptual abilities is described

by recently very influential theory of predictive coding, that can also give an explanation of SoA (Friston, 2012).

It has been discussed whether alteration in SoA is present only in patients with first-rank symptoms or it might be related to the disease itself, i.e. also in patients without those symptoms (Farrer & Franck, 2007; Moore & Bravin, 2015).

1.1 Goals

In our study we wanted to test, whether:

- First-episode schizophrenia patients have diminished capacity and accuracy to continuously distinguish self-produced from other-influenced movement, and thus reduced SoA measured by GCT.

We expected attenuated performance in SoA task in the patient group.

- Patients' performance is modulated by presence of first rank symptom and how their performance relates to the dimensions of psychopathology in PANSS sub-scales.

We expected that patients who experienced first-rank symptoms would be less successful in SoA task, their performance would negatively correlate with Positive PANSS but there would be no correlation with other sub-scales.

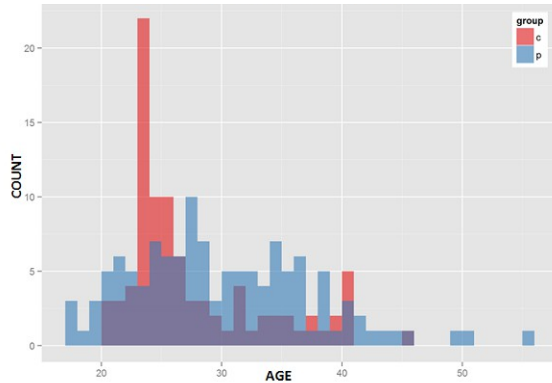
2 Methods

2.1 Subjects

Patients group consists of 122 participants diagnosed with schizophrenia (64 males) for the first time. The eligibility criteria for the patients were age between 18-60 years, diagnosis of psychotic disorder according to ICD 10, absence of organic mental illness, mental retardation, or serious neurological illness. Patients were invited to participate in this study during their hospitalization in Psychiatric Hospital in Bohnice.

Control group consists of 90 healthy subjects (43 males) with no history of psychiatric disease in their

close family. The eligibility criteria for the healthy controls were age between 18-60 years, absence of mental illness, absence of psychiatric disease in close family (in first-degree family members), absence of mental retardation and serious neurological illness. Participants were invited to participate in this study by experimenters. Age structure of participants is depicted in Fig. 1.



Obr. 1: AGE HISTOGRAM: patients (blue) vs. controls (red)

2.2 Data gathering

Data of healthy controls started being collected in 2013. All participants were given basic information about the study (including purpose of the study, brief information about experimental task, etc.). Informed consents from participants from both groups were obtained before they entered the study. They completed the Ghost city task in MRI scanner. The patients were assessed in PANSS in the days near to testing.

2.3 Methods

The overview of assessments in all participants is presented in the Tab. 1. The first part of the analysis, that compares both groups, takes into account only the first assessment. The following analysis is of longitudinal character and tests the individual performance across time and - in patients - also with regards to the presence of symptoms.

| | Assessments | |
|----------------------------|-------------|----|
| | FES | HC |
| Ghost city | X | X |
| PANSS | X | |
| First rank symptoms | X | |

SS

Positive and Negative Syndrome Scale (Kay, Fiszbein, & Opler, 1987) is an instrument used for measuring variety of symptoms in schizophrenia. The scale is divided into three subscales: Positive scale (consisting

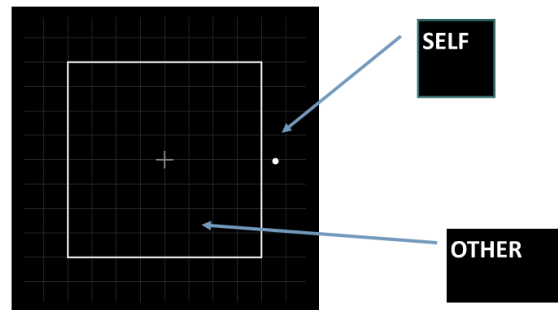
items such as delusions, hallucinatory behavior or suspiciousness), Negative scale (e.g. lack of spontaneity or emotional withdrawal) and General psychopathology scale (e.g. anxiety, tension, uncooperativeness), consisting in 30 items in total, each of which is scored in severity on scale from 1 (absent) to 7 (extreme).

2.3.2 Schneider first-rank symptoms

Psychiatrists who interviewed the patients using PANSS gave also score: 1 for presence of FRS during lifetime or 0, if FRS had not been present.

2.3.3 Ghost city task

Ghost city is a motor task with distorted visual feedback. Task environment is defined by a grind with a square (box) in the centre. The square splits whole area into two parts – peripheral area (outside the box) and central area (inside the box). Participants' objective is to decide whether the movement of a dot presented on the screen is their own (SELF condition) or influenced by experimenter (OTHER condition). If the movement is detected as self-produced, they should move the dot outside the box. On the contrary, if participants deem the movement of the dot to be influenced by experimenter, they should move the dot inside the box.



Obr. 2: Play field, in the SELF condition the dot should be moved outside, while in the OTHER condition inside the square

The participants are told that from time to time another experimenter is sitting by a computer using a similar joystick, with which he will influence the participant's movement projected on the screen. In fact, the movement is not influenced by a random intrusion of an experimenter, but is periodically distorted on a basis of an algorithm that changes the displayed movement vectorially (with varying length and direction of distortion vector). The period of the SELF condition and the OTHER condition is repeated 12 times during the whole session, with 20 sec duration of each condition, so the whole task lasts less than 10 minutes.

Tab. 1: Overview of Assessments

2.3.1 PAN

Participants control a white dot by moving a joystick placed in their hand. Because the testing was done in MRI scanner, participants were using a fMRI-compatible joystick placed in their hand while they watching a screen above their head reflected in mirror glasses.

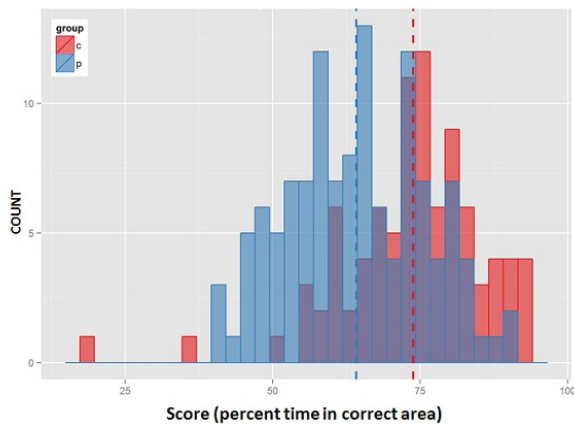
3 Results

3.1 Between group comparison

To test whether there is an alteration in Sense of Agency in patients compared to healthy controls (measured as GCT score) we used two-sample t-test.

3.1.1 Performance in Sense of Agency task

Comparison of patient and control group shows statistically significant difference in performance between both groups ($t(189) = -5.83, p < .001$). Histogram of performances of both groups is in Fig. 3.



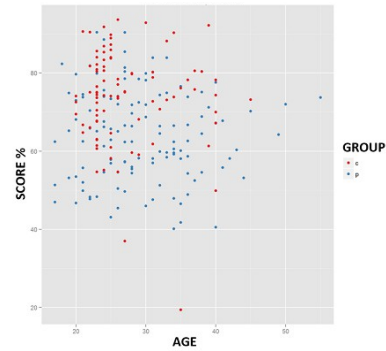
Obr. 3: Histogram of performance in both groups

3.1.2 Influence of age on performance

In the histogram in Obr. 1 we can see there is a difference between patient and control group in age. Thus, we did an analysis to test, whether age might correlate with performance, which could suggest there was a bias in our data samples that could influence the results. We tested that, and our results show this should not be the case (Fig.4).

3.2 Patients' performance

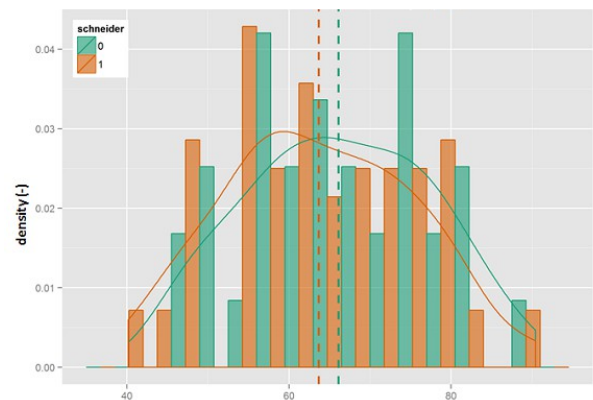
In the patient group we analysed how clinical variables relate to their performance. We tested whether SoA (measured as GCT performance) correlates with the presence of first-rank symptoms or with PANSS subscales.



Obr. 4: Age versus score in GCT

3.2.1 Sense of Agency and first-rank symptoms

We expected that deficit in SoA will more pronounced in patients with first-rank symptoms. However, our data doesn't support this hypothesis (Fig.5). Two-sample t-test results are: $t(65) = -1.05, p = 0.30$.



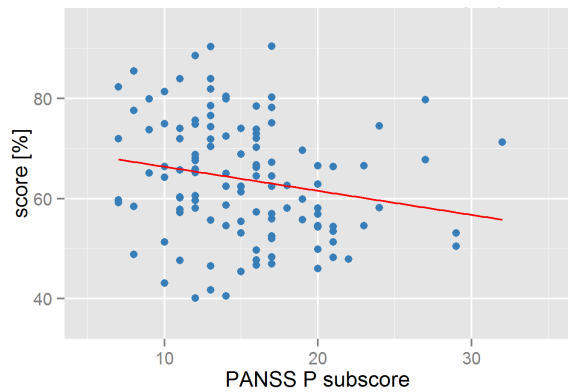
Obr. 5: Histogram of patients' performance in relation to the presence of first-rank symptoms (1=present)

3.2.2 Sense of Agency and PANSS subscales

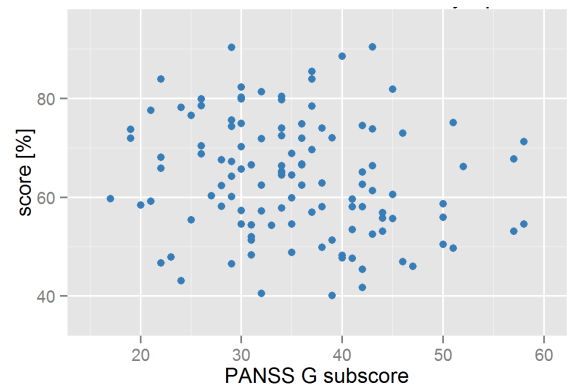
We tested the relation of patients' performance in the Ghost City Task to four PANSS scores - positive, negative, global and total. After correction for multiple testing the Alpha-level should be .0125.

Positive symptoms

There is a negative correlation between the score in Positive Symptoms subscale and performance. The results of linear regression are: $\beta_1 = -.48, p = .03, R^2 = .04, F(1, 116) = 4.86$ (Fig.6).



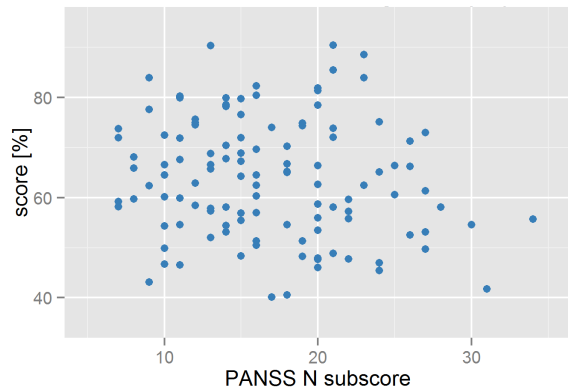
Obt. 6: Correlation between Positive PANSS subscore and performance score in GCT



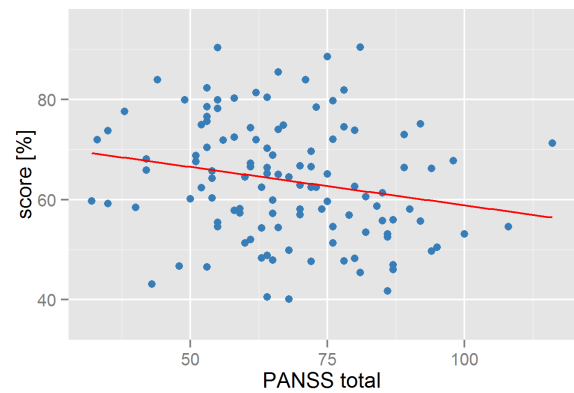
Obt. 8: PANSS Global subscore and performance in GCT

Negative symptoms

Test of correlation between Negative PANSS score and performance in GCT didn't yield to a result that would suggest a relation between both scores. Results of linear regression are: $\beta_1 = -.29$, $p = .11$, $R^2 = .04$, $F(1, 116) = 2.51$ (Fig.7).



Obt. 7: Negative PANSS subscore and performance in GCT



Obt. 9: PANSS Total and performance in GCT

Global symptoms

Also, testing a relation between Global PANSS subscore and GCT doesn't suggest any dependence. Linear regression results are: $\beta_1 = -.23$, $p = .05$, $R^2 = .03$, $F(1, 116) = 3.68$ (Fig.8).

PANSS Total

When taking into consideration PANSS the sum of sub-scores, the results in regression analysis yields to results: $\beta_1 = -.15$, $p = .02$, $R^2 = .04$, $F(1, 116) = 5.36$ (Fig.9).

Summary

In sum, although Positive and Total PANSS is negatively correlated with GCT score, these relations are not statistically significant after Bonferroni correction for multiple testing. Total PANSS is comprised also of Positive PANSS, i.e. the only subscale that shows a relation to GCT, so we infer that Total PANSS correlation with GCT might actually reflect similar relation of Positive PANSS and GCT. Although statistically nonsignificant, the results show a tendency for negative correlation between Ghost City Task score and presence of Positive Symptoms as captured by PANSS.

4 Discussion and follow-up study

We found a statistically significant difference between healthy controls and schizophrenia patients in Ghost City Task; patients' performance was attenuated, and that was not influenced by the difference in age between both groups. These results would suggest that the problem in differentiation between self and other is possible to register also on the level of distinguishing self-initiated from other-influenced movement.

Another interesting finding is that there was a tendency for a negative correlation between performance in Ghost City task and Positive PANSS score (connected also to Total PANSS score). This result suggests that presence of positive symptoms, such as delusions and hallucinations, corresponds, indeed, to the diminished performance in our Sense of Agency task and thus is possible to observe on the level of behavioural motor test. The more pronounced positive symptoms, the worse was the ability to discriminate self- from other-produced actions, albeit this relation should be considered more cautiously as a tendency.

Schneider's symptoms did not modulate the impairment of performance in our SoA task; both patients with and without first-rank symptoms were successful in decisions about ownership of movement to a similar degree. This may suggest that impaired SoA could be present also in patients without Schneider's symptoms and thus not related exclusively to first-rank symptoms. Such results may seem a bit contradictory, because Schneider first-rank symptoms and Positive PANSS overlap considerably. Importantly, scores of first-rank symptoms were based on their presence during life-time, on the other hand PANSS was scored based on concurrent presence of the symptoms at the time of - but not directly during - testing.

Could there be other explanations for our finding of diminished SoA in the patient group? Patients might have had worsened performance due to confounding factors when using MRI in schizophrenia patients - we can speculate that their performance might have been influenced by more intense stress experienced during testing or by fluctuation in attention. All that might lead to not following the task. On the other hand, the patients' performance was related neither to the Global scale nor to the Negative scale, so their score should not be influenced by poor attention, apathy or uncooperativeness. Nevertheless, those measures are indirect.

Some studies show that in FES there might present a motor deficit, which might also account of patients' attenuated performance (Bilder et al., 2000).

For the reasons mentioned above, we developed a control task - a brief motor task very similar to the experimental one. In the same test environment (white square on a black background and a dot controlled by joystick) participants move the dot inside or outside the square based on its colour, which changes pseudo-

randomly (again algorithm-based) between green and red. This way we can take into account also participants' psychomotor tempo - the reaction time between noticing of a change in colour of the dot and moving into the correct area - and their ability to control the joystick. Undergoing the control task could also be considered as a type of training for the experimental task, since the only change between both tasks is the criteria on which participants decide where to move the dot, i.e. a colour or expected position and trajectory.

We have started to retest participants also after some time (approximately one year) to see how their performance changes. In patients we are especially interested in the relation of SoA and expected changes in PANSS (related to remission) to examine how mutually dependent those domains are.

Another direction of our research is to include healthy sibling of schizophrenia patients into the study. Possible alterations in SoA in siblings would suggest that SoA disturbance might be connected not only to schizophrenia itself, but can be an endophenotype of the disease and thus have hereditary underpinnings.

Acknowledgement

This work was supported by the project „National Institute of Mental Health (NIMH-CZ)“, under grant number ED2.1.00/03.0078, and the European Regional Development Fund, and by the project “Sustainability for the National Institute of Mental Health”, under grant number LO1611, with a financial support from the Ministry of Education, Youth and Sports of the Czech Republic under the NPU I program and by the project GA UK number 94215.

References

- Bilder, R. M., Goldman, R. S., Robinson, D., Reiter, G., Bell, L., Bates, J. A., ... others. (2000). Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *American Journal of Psychiatry*, 157(4), 549–559.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, 6(6), 237–242.
- Farrer, C., & Franck, N. (2007). Self-monitoring in schizophrenia. *Current Psychiatry Reviews*, 3(4), 243–251.
- Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology*, 83(2), 248–252.
- Frith, C. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Psychology Press.

- Frith, C. (2012). Explaining delusions of control: The comparator model 20years on. *Consciousness and Cognition*, 21(1), 52–54.
- Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, 13(2), 261–276.
- Moore, J. W., & Bravin, J. (2015). Schizotypy and awareness of intention: Variability of W judgments predicts schizotypy scores. *Psychology of Consciousness: Theory, Research, and Practice*, 2(3), 283.
- Sass, L. A. (2014). Self-disturbance and schizophrenia: Structure, specificity, pathogenesis (Current issues, New directions). *Schizophrenia Research*, 152(1), 5–11.
- Sass, L. A., & Parnas, J. (2003). Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin*, 29(3), 427–444.
- Strauss, J. S., & Carpenter, W. T. (2013). Assessment and Diagnosis. In *Schizophrenia* (pp. 19–42). Springer Science & Business Media.
- Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, 133(1), 262–271.
- Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain*, 133(10), 3104–3112.

Porozumenie literatúre

Jana Kuzmíková

Ústav slovenskej literatúry SAV
Konventná 13, 811 03 Bratislava
jana.kuzmikova@gmail.com

Abstrakt

V príspevku upozorňujem na dva rozdielne spôsoby recepcie literatúry: spontánne a odborné čítanie. Napriek dokumentovaným (fMRI) rozdielom sa na oboch typoch čítania podieľajú pocity a spomienky čitateľa. Na citové reakcie (stavy) sa následne viažu (nové) významy čítaného textu. Mojmím zámerom je naznačiť, ako vzniká porozumenie literatúre v procese čítania u každého čitateľa umeleckých textov, pričom bokom ponechávam konvenčnejší prístup, ktorý chápe literárne porozumenie ako výsledok (odborných) interpretácií, zohľadňujúcich kultúrno-historické kontexty textu.

1 Empirické výskumy kognitívnej literárnej vedy

V 80. rokoch 20. storočia si podľa príkladu kognitívnej lingvistiky aj niektorí literárni vedci osvojili požiadavku, že v humanitnom výskume je potrebné rešpektovať vtelesnenosť (embodiment) ľudskej mysle. Začala sa formovať kognitívna literárna veda. Medzi inými jej cieľmi je dôležitý aj zámer, že je žiadúce rámcovať a podložiť interpretácie literárneho textu psycho-fyziologickým výskumom procesov mysle. Na realizácii tohto zámeru sa podieľajú aj empirické výskumy literárnej komunikácie.

Z empirických pozorovaní vieme, že bežného čitateľa nezaujímajú, ako má text interpretovať, ale ak číta, tak predovšetkým z dôvodu, že prítomné čítanie má pre neho istý aktuálny význam. V projektoch kognitívnej literárnej vedy preto nie je prvoradá interpretácia textu, ale výskum sa zameriava na širšiu skúsenosť z literatúry. Do procesu produkcie aj porozumenia literárnemu textu sa zapájajú mentálne mechanizmy vnímania, usudzovania, vyvodzovania aj cítenia.

Ukazuje sa, že laickí čitatelia si vnímavo všimajú (hoci odborne nerozoberajú) fonetické, syntaktické aj sémantické zvláštnosti umeleckých textov. Ale príťahuje ich istý druh (estetickéj) skúsenosti, informačný obsah výpovede, príbeh, uvažujú o činoch a osudoch postáv, sú zasiahnutí emóciami a reflektujú ich. V experimentálnych výskumoch temer nikto nezačal predložený umelecký text samoučelne

interpretovať bez viazanosti na určité osobné skúsenosti a záujmy, ktoré v danom texte (žánri) „hľadal“ či nachádzal (Miall, 2007, 26). Čitatelia mobilizujú svoje osobné skúsenosti a spomienky, aby dokázali umelecký text vnímať v potrebných súvislostiach. Iste, aj neliterárne texty vyvolávajú u čitateľa isté spomienky, ale literárne texty mu dvakrát častejšie pripomínajú jeho vlastné spomienky, teda také, ktoré sa vyznačujú individuálnym hľadiskom spätým soznačením o mieste, čase a zdroji udalosti/skúsenosti. Neliterárne texty vyvolávajú spomienky sprostredkované najmä nevlastnými zdrojmi (čo niekto iný povedal, napísal a pod.) (Sielman a Larsen, 1989, 174).

Vnímanie literárnych textov oproti neumeleckým je teda odlišné. Rozdiely sú však aj v recepcných procesoch v rámci samotnej beletrie. Intuitívne by sme predpokladali, že pozorné, resp. školené čítanie literatúry stavia na procesoch, ktoré sa zapájajú pri bežnom čítaní pre radosť, keď je čitateľ vtiahnutý do deja a pohľtený osudmi obľúbených postáv. Výskum N. M. Phillipsovej (2015) však dospel k prekvapujúcemu zisteniu. Literárnovedné čítanie neprehľbuje, nestavia na schopnostiach a procesoch bežného čítania. Oba druhy čítania síce vyvolávajú v mozgu rozsiahlu aktivitu, ale aktivované oblasti sú rozdielne, neprekrývajú sa.¹ To znamená, že čítanie, pri ktorom si zvlášť všimame napríklad jedinečné formálne črty umeleckých textov a interpretujeme ich, si vyžaduje úplne iné kognitívne schopnosti ako spontánne čítanie. A ešte inak: aj keď v recepcii môžeme prepínať (a väčšinou aj prepíname) medzi čítaním pre radosť a školeným čítaním zacieleným na interpretačný rozbor textu (sujetu, prostredia, hľadiska rozprávača, poetiky, štýlu rozprávania, obraznosti a pod.), treba obrátiť primeranú výskumnú a vyučovaciu pozornosť aj na dlhodobu podceňované čítanie pre radosť a oddych, pretože aj tento spôsob recepcie má svoje kognitívne parametre a rozsahy, ktoré nie je vhodné odsúvať a redukovať. Bežné čítanie fikčných príbehov napríklad bezprostredne zlepšuje empatiu, interpersonálnu percepciu a sociálne schopnosti nezávisle od

¹ Pri pozornom, odbornom čítaní sa aktivizuje napríklad aj somatosenzorický a motorický kortex, pri spontánnom čítaní pre radosť sa aktivizuje napríklad temporálny kortex (Phillips, 2016, 61-62).

individuálnych rozdielov recipientov (Mar a kol., 2008).

Pri plynulom čítaní, v ktorom sa strieda spontánne čítanie s pozorným, sa podnecujú značne rozsiahle procesy v ľudskej myslí (mozgu), čo samo osebe je argumentom na podporovanie literárnych a čitateľských aktivít. Literárna recepcia obojakého druhu poskytuje rôzne potrebné kognitívne stimulácie, lebo teória mysle nie je jedným univerzálnym súborom inferencií pre nalinkované udalosti a situácie, ale je tvorená mnohými súbormi adaptácií.

2 Čítanie v literárnej komunikácii

Napriek tomu, že čítanie pre radosť a „pozorné“, školené čítanie sa prejavujú netotožne cez neprekrývajúce sa mozgové (neurónové) mapy, na oboch spôsoboch čítania sa zúčastňujú pocity a city. Čítanie je preto aktuálne zdôrazňovaným článkom literárnej komunikácie (najmä čítanie na strane čitateľa, ale aj autora). Bol napríklad pozorovaný identifikovateľný rozdiel medzi vnímaním a pociťovým stotožnením sa čitateľa s postavou oproti vnímaniu fiktívneho deja, prostredia a času, lebo tieto kategórie viac podliehajú štruktúrálnej analýze a sekvenčnej logike (bližšie Kuzmíková, 2015).

Čítanie je v recepčných súvislostiach definovateľné ako subjektívna skúsenosť bez zjavných znakov a podnetov konania, ale s operatívnou funkciou, ktorá produkuje súdy, preferencie a pod. Je to vnútorný, subjektívny výklad aktuálneho stavu, v akom sa nachádza emočný motivačný systém (bližšie Hvorecký, 2011).

2.1 Citové prototypy a prototypové príbehy

Už na základe bežnej skúsenosti možno tvrdiť, že akákoľvek citová udalosť v určitom okamihu sa vždy istým spôsobom vzťahuje k absolútnej prototypovej emócii, ktorou je šťastie (alebo jeho opak smútok). Šťastie môže byť fyzické, osobné, sociálne alebo transcendentálne a etické.

Príbehom o duchovnom šťastí je harmonické stretnutie sa dvoch či viacerých ľudí, napríklad uzmierenie sa predtým rozdelenej rodiny, ale najvýznačnejší príbehový model šťastia predstavuje milenecký romantický príbeh. Je to prototypový príbeh jednoty v láske a zároveň najčastejší, základný naratívny model. Druhým základným prototypovým rozprávaním je hrdinský príbeh, v ktorom hrdinovia bojujú za spoločenské postavenie a politickú moc. Tretím druhom prototypového príbehu je príbeh obetovania, ktorého cieľom je prežitie protagonistov, zabezpečenie ich fyzických potrieb (jedlo, pitie). Z kognitívneho hľadiska treba zobrať do úvahy, že v prototypových príbehoch je výsledný pocit šťastia dosahovaný v súhre viacerých prototypových pocitov a citov, ktorými sú

najmä zaľúbenosť, slasť, hnev, strach a hlad (= túžba jesť, piť) (Hogan, 2003).²

2.2 Čítanie v procese čítania

Z prototypových po-citov človeka sa teda odvíjajú základné rozprávané príbehy. Okrem týchto abstrahovaných citovo-textových príbehových schém (primárnych sujetov) sa v rámci naratívneho rozprávania čoraz viac študuje aj čítanie čitateľa v aktuálnom procese čítania (Kuzmíková, 2011). Nijaká štruktúra sujetových vzťahov či sieť inferencií (odvození) nedokáže zachytiť „neobmedzenosť“ textu v jeho všemožných čitateľských realizáciách počas nepretržitého vývinu textových významov v čitateľovej myslí. V tejto neurčitosti však môže pomôcť čítanie človeka ako princíp, ktorý vedie proces porozumenia textu. Pre toto stanovisko jestvujú predpoklady (Miall, 2007, 48-55):

1. Čítanie je sebareferenčné a preto umožňuje, aby čitateľ aplikoval v procese porozumenia textu skúsenosti a hodnoty, obsiahnuté vo vlastnom koncepte svojho Ja. Sebareferencia pritom nemusí prebiehať sebauvedomelo, vedome.
2. Čítanie umožňuje kategorizáciu textových elementov naprieč doménami (napr. motivickými plánmi prostredia, deja a postáv).
3. Čítanie má anticipačnú kognitívnu funkciu, lebo „nastavuje“ štruktúru porozumenia textu v kontexte čitateľovej percepcie, pamäti, poznania.

3 Konkretizácia čítania v recepcii: Vilikovského Pán spomienok

3.1 Sebareferencia

S čítaním čitateľa zručne pracuje Pavel Vilikovský v poviedke Pán spomienok zo zbierky Čarovný papagáj a iné gýče (2005). Na začiatku poviedky buduje autor jednotlivé odseky ako krátke epizódy, ktoré sa končia istým mentálnym podnetom:

„Že v prázdnej miestnosti znejú hlasy ináč, je známe, a dá sa to ľahko vysvetliť: zmenená akustika. Ale keď som sa pozrel von oknom, prekvapilo ma, že z prázdnej miestnosti je aj iný výhľad, cudzí.“ (Vilikovský, 2005, 337)

Ladenie úvodného odseku poviedky je pre čitateľa negatívne, evokované slovami prázdny, cudzí. Rozprávačov odcudzený výhľad sa potvrdzuje tým, že nedokáže uzrieť otcov hrob: „Teraz sa mi to nedarilo“ (337).

Mentálny podnet sa stupňuje na začiatku nasledujúcej epizódy. Medzi uvádzajúcim prvým a nasledujúcim druhým odsekom vnímame negatívny pocit z istého druhu túžby, čo sa nedarí a v ceste jej stojí prekážka

² Viac o základných emóciách, teóriách emócií a emóciách v literatúre Gáfrík (2014).

(kvôli oblakom nemôže rozprávač dovidieť na cintorín s otcovým hrobom). Záporný stav z konca prvého odseku (neuskutočiteľné prepojenie živého s mŕtvym človekom), keďže je predstavený na citovej úrovni, neurčuje presne, čo bude nasledovať.

Prestrih do druhého odseku začína: „„No, a na záver nám tu ostal najťažší oriešok,“ povedal brat od kachiel.“ (337) Zmena externého pohľadu na pohľad do interiéru ešte zosilňuje nastolenú negatívnu situáciu, prežívanú rozprávačom aj cez jeho „tichú zúrivosť“.

Pri „zúrivo“ vyprázdňovaní maminho (rodičovského) bytu si rozprávač zoberie len otcovu fotografiu a plechové pľuvadlo. Zvažuje, že pľuvadlo sa dá použiť ako popolník. Úvaha o zámene pľuvadla na popolník vyvoláva nielen elementárne asociácie spopólnenia spojeného so smrťou. Za funkčnou zmenou danej veci je aj isté rozhodnutie rozprávača. A opäť sa zopakuje forsírujúci autorský postup, uplatnený už o odsek vyššie, keď v prvej vete nasledujúceho, tretieho odseku čítame: „Musíme sa rozhodnúť, čo urobíme s týmto,“ povedal brat a ukázal hlavou na kôpku písomnosti“ (337). Vysvitne, že tým ťažkým orieškom na rozhodnutie je spálenie otcovej písomnej pozostalosti.

Naznačila som podnecujúce faktory čitateľovho precitovania spisovateľovej výpovede. Pokiaľ ide o následné recepčné procesy, pocity jednak vyvolávajú isté mentálne schémy, scenáre (napríklad citové prototypové príbehy), ale môžu spôsobiť aj ich narušenia, multivalentné skríženia, deviácie. Okrem toho, po-city nie sú späté len s ustálenými schémami, lebo aj jednotlivé slovné spojenia (napríklad rozprávačovo prirovnanie jeho erotických lások k drevenej latrine) alebo slová (pľuvadlo) môžu vniknúť do citového naladenia čitateľa.

V úvodnej časti poviedky z môjho sebareferenčného pohľadu sú citovo najintenzívnejšie tri výroky/predstavy: o cudzom výhľade, ktorý nedovídi na otcov hrob; zmienka o tichej zúrivosti rozprávača a „záhadný predmet“ pľuvadla, do ktorého sa nepľulo, ale napriek tomu či práve preto zastupuje zaniknutú domácnosť. Všetky tri zvýraznené motívy majú negatívnu citovú valenciu. Ich opačný, potlačený potenciál je pozitívny a časom naozaj dochádza k (ambivalentnému) prehodnoteniu pocitovej akosti uvedených podnetov, čo má zásadný dosah aj na porozumenie: napriek prvotnej zúrivosti, aby všetko z bytu zmizlo preč, si rozprávač predsa len ponechá časť otcovej písomnej pozostalosti a napokon aj obrazne „dovídi“ až do otcovho hrobu.

3.2 Cítanie naprieč doménami

Cítanie je produktívne v recepčných procesoch aj takým spôsobom, že prechádza naprieč štruktúrnymi textovými aj pojmovými doménami. Vďaka tomu pôvodne zaznamenané významy a schémy jednak získavajú latentne tušené, resp. komplementárne „zabudované“ nové citové zafarbenia a jednak cítanie

prestupuje za a naprieč doménami (napríklad pľuvadlo z motivického plánu prostredia vyvoláva okrem znechutenia aj zvedavosť ako „záhadný predmet“ a ako taký naznačuje aj isté rodinné pomery, správanie sa postáv a súvislosť medzi rozprávačom a jeho otcom, oboma tuhými fajčiarmi: čiže pľuvadlo-popolník ako spúšťač istých pocitov funguje aj v motivickom pláne postáv a deja). Pretože rozprávač neustále narúša čitateľom osvojené významy a jeho schematické očakávania, čitateľ do konštruovania významov a reprezentácií naratívu automaticky zapája alternatívny spôsob porozumenia, ktoré sa formuje cez pocitované účinky z rozprávania. Túto tézu bližšie rozvádzam v nasledujúcej podkapitole.

3.3 Anticipačná funkcia cítania

Vyvolané pocity môžu pretrvávajúť v mysli (pracovnej pamäti) dlhšie než ich konkrétne spúšťače, rezonujú v recepcii ako jej „spodný tón“ a zároveň alternatíva momentálneho cítania a porozumenia rozprávania v danom okamihu. Platí to aj pre exponovaný motív pľuvadla. Jeho citovú (ambi)valenciu umocňuje fakt, že napriek tomu, že je zástupným symbolom (zaniknutej domácnosti), má nejasný kognitívny obsah. Jeho funkčnosť, zmysel je neurčitý, následkom čoho sa otvára jeho metaforický potenciál. Predmet najprv evokuje negatívne pocity pľuvania (odpor, znechutenie), potom možnú zmenu na popolník s pridruženými negatívnymi konotáciami spopólnenia (smrti) a škodlivého fajčenia (choroby) (prípadne sa u čitateľa fajčiara môžu vynoriť komplementárne pocity príjemného fajčenia). V uvedenej škále sa funkčné vlastnosti predmetu ešte nevyčerpávajú, naopak, odкрýva sa jeho myslený metaforický rozsah, vďaka čomu sa v súhre s pocitmi anticipujú viaceré potenciálne naratívne línie. V tomto zmysle cítanie predznamenáva a predštruktúruje význam rozprávania ako celku. Pojem pľuvadla-popolníka vo svojom prototypovom jadre obsahuje predstavu nádoby, schránky, akou je aj koryto, v ktorom sa rozprávač ako trojročný chlapec počas vojny zabával vyspevovaním a jeho otec sa tomu tešil. V tomto mieste výpovede recepčne nastáva pocitová zmena. Završuje sa predstavou „hlineného koryta“, hrobu, v ktorom leží rozprávačov otec. Rozprávač si predstavuje, že otec si v hrobe možno aj spieva, čo vyvoláva radostnejší pocit aj u čitateľa. Vďaka rozprávačovej šťastnej detskej spomienke sa od pľuvadla, vzbudzujúceho skôr nepríjemné pocity, nakoniec dočítame k úsmevnému záveru. Je však lokalizovaný v hrobe, takže zmiešaný kladný, ale aj melancholický pocitový záver vníma empatický čitateľ v celom spektre precitovania Vilikovského silno ironického rozprávania, v ktorom je materiálne, telesné a skutočné nahradené spomienkou a ilúziou.

Ak treba ešte názornejšie podprieť recepčnú skúsenosť, že cítanie predznamenáva porozumenie, môžem pripomenúť rozprávačovú zmienku, že on si

lásky nikdy nevyberal, na neho „sa jednoducho zvalili ako drevená latrina“ (352). Predpokladám, že u čitateľa vyvolá dané prirovnanie istú citovú odozvu, ale v celom príbehu nenájdeme priamejšie objasnenie ani návod, ako ho chápať. Čitateľ reaguje pocitovo (môže byť nepríjemne prekvapený, pociťovať trápnosť alebo hoci byť aj pobavený, atď.), ale ak má byť dané prirovnanie aj interpretované, tak hlavne prostredníctvom anticipačnej funkcie čítania, ktorá sa však v rámci tejto poviedky nemôže spoliehať na kognitívne konkrétnejšie textové indície. Nápomocné sú tu len mimotextové zdroje – osoba čitateľa, jeho čítanie a domýšľanie. Štandardné predstavy spojené s pojmom lásky (porovnaj romantický prototypový príbeh), čiže schémy aplikované na konceptuálnej kognitívnej úrovni sa pri uvedenom motíve rušia, znefunkčujú, čo je tiež sprevádzané istými pocitmi. Keď má čitateľ k dispozícii len svoje čítanie, ono spúšťa proces rekontextualizácie, tvorby novej schémy s novými významami. Z neurofyziologického hľadiska vysvetľuje tento proces D. Miall (2007, 147-49): Hlavnú úlohu zohráva nerovnaký priming v ľavej a pravej hemisfére. PH, spojená s limbickým systémom, má dlhší čas vyvolania reprezentácií a asociácií, vďaka čomu sa zväčšuje ich relačná vzdialenosť a metaforický potenciál. Keďže spojenia v PH sú informačne odľahlejšie, s podporou limbického systému (cítienia) napomáhajú nové schémy a riešenia.³

Čítanie teda nielen sprevádza kognitívne procesy (ako napríklad hodnotenie), nielen koordinuje prepínanie medzi kognitívnymi plánmi (Oatley a Johnson-Laird, 1987), ale má aj významnú kognitívnu funkciu, lebo anticipuje modifikáciu zavedených, „zabehnutých“ mentálnych schém s im náležiacou pociťovanou valenciou a podnecuje nové mentálne schémy (s možnou lepšou adaptačnou funkciou). Experimentálne je potvrdené, že čítanie je platformou na zaznamenávanie významných obsahov a čít udalostí, rozvíjaných v priebehu rozprávania, pričom prekračuje obmedzenia pracovnej pamäte. Zmeny medzi naratívnyimi situáciami signalizuje zvyšovanie alebo znižovanie pocitov napätia. Čítanie je „prepravným prostriedkom“ pre empatiu, sympatiu aj odmietnutie a slúži na zmenu postojov voči objektu rozprávania. Zdá sa, že čítanie má výraznejšiu recepčnú funkciu (silu) než rozvíjanie deja, point of view a empatická

³ Okrem toho je však z hľadiska čítania dôležité, že vysoko emočné slová, kladné aj záporné, aktivujú širokú mozgovú sieť, ktorá zahŕňa ľavý DLPFK (dorsolaterálny prefrontálny kortex), bilaterálny mediálny prefrontálny kortex, amygdalu, gyrus cinguli, inzulárny kortex a precuneus. Zvýšená činnosť ľavej insuly sprevádza zvyšujúce sa príjemné stimuly, zatiaľ čo zvýšená činnosť v pravom DLPFK a pravom precuneu sprevádza zvyšujúce sa averzívne stimuly (Phillips, 2016, 68).

spriaznenosť s postavami, pretože čítanie zahŕňa všetky spomenuté komponenty (Miall, 2007, 137).

Pod'akovanie

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

Gáfrík, R. (2014) Emocionálne pôsobenie literatúry - znázornené na príklade Kafkovej poviedky Ortieľ. V kol. monog. Literatúra v kognitívnych súvislostiach, str. 111-126.

Hogan, P. C. (2003). The Mind and Its Stories. Cambridge University Press, 1. vyd.

Hvorecký, J. (2011). Teórie emócií a fikcia. World Literature Studies. 3 (3): 53-61.

Kuzmíková, J. (2011). Pokus o kognitívne čítanie Vilikovského románu Posledný kôň Pompejí. Svět literatury. XXI (44): 37-58.

Kuzmíková, J. (2015) Recepčia literatúry z kognitívnovedného pohľadu. World Literature Studies. 7 (4): 36-51.

Mar, R. A.; Oatley, K.. a Djikic, M. (2008) Effects of Reading on Knowledge, Social Abilities, and Selfhood. Theory and Empirical Studies. V kol. monog. Directions in Empirical Literary Studies. John Benjamins, str. 127-137.

Miall, D. S. (2007). Literary Reading. Peter Lang, 2. vyd.

Oatley, K. a Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. Cognition and emotion. 1: 29-50.

Phillips, N. M. (2015) Literary Neuroscience and History of Mind: An Interdisciplinary fMRI Study of Attention and Jane Austen. V kol. monog. The Oxford Handbook of Cognitive Literary Studies, str. 55-81.

Sielman U. a Larsen, S. F. (1989). Personal resonance to literature. Poetics. 18: 165-177.

Vilikovský, P. (2005). Prózy. Kalligram, str. 337-355.

Modelling of the Reference Frame of the Ventriloquism Aftereffect

Ing. Peter Lokša, doc. Ing. Norbert Kopčo PhD.

Authors' addresses respectively: Institute of Computer Science, Faculty of Science, Šafárik University in Kosice
peter.loksa@gmail.com, norbert.kopco@upjs.sk

Abstract

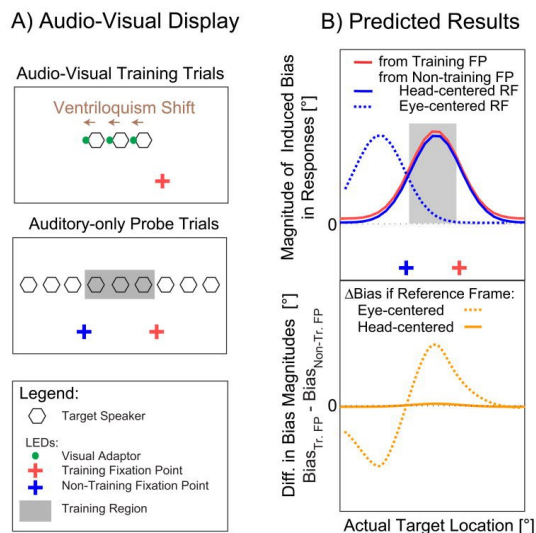
The human brain extracts information from various senses in order to represent the physical space. To integrate spatial information from the visual and auditory modalities, the modalities need to be aligned as each of them represents spatial information in a different reference frame (RF). The visual RF is aligned with the direction of eye gaze while the auditory one is aligned with the orientation of the head. And these two ways of such alignment of audio-visual spatial RF are competitively hypothesized. But two main parts of experimental data are inconsistent in suggestion of spatial RF. In this article there is the result of the partial modelling of experimental data with aim of resolving such inconsistency.

1 Introduction

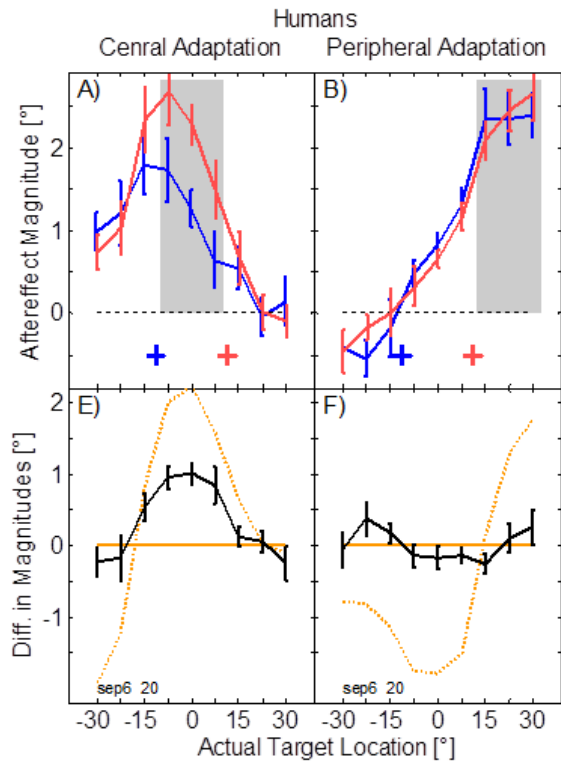
Vision plays an important role in how the brain processes auditory information (Alais, Burr, (2004)). In the spatial domain, vision provides guiding signals for calibration of spatial auditory processing. This can be illustrated by the ventriloquism aftereffect illusion in which repeated pairings of 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)). The current study models the ventriloquism aftereffect to elucidate several aspects of the mechanisms underlying visual calibration of auditory space and to characterize its spatial properties of the calibration.

2 Early study (Kopčo N (2009))

The experiment was done to investigate reference frame of ventriloquism aftereffect (Kopčo, (2009)). There were 7 humans participating in this experiment as subjects. Using responses from this experiment the aftereffect magnitude variable was obtained in order to identify the questioned reference frame (RF), which can be eye-centered or head-centered.

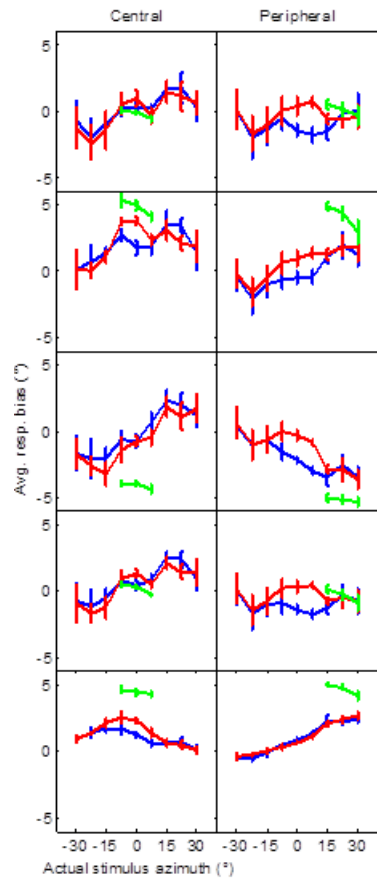


Obr. 1: A) Symbols in "Audio-Visual Training Trials" panel mark the azimuths of stimuli provided to subjects in so called audiovisual training trials, in the way that the azimuthal relative shift between physical location of stimuli, that are synchronous, are constant within given experimental session for each session. The symbols in "Auditory-only Probe Trials" panel mark azimuths of so called auditory-only trials, which are 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.

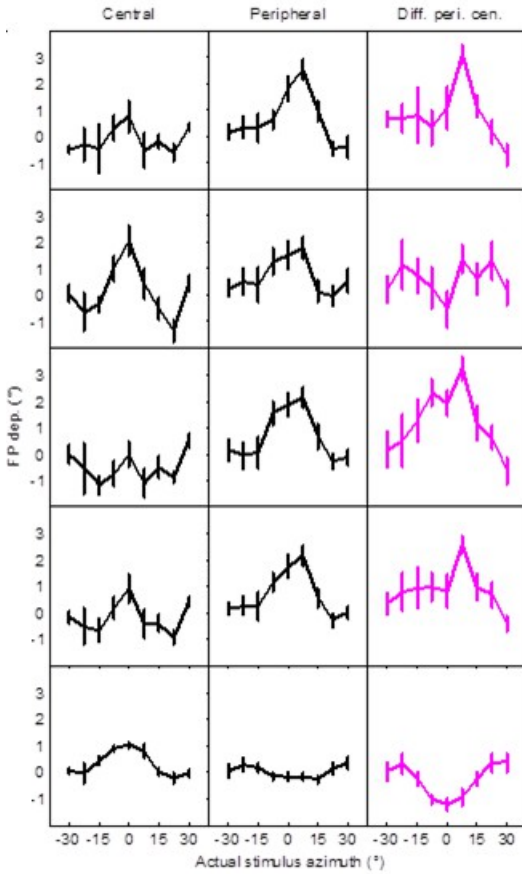


Obr. 2: Magnitude of ventriloquism aftereffect and reference frame determination according to difference between training vs. non-training trials. Red/Blue line - separation of probe auditory-only trials according to the pre-trial eye gaze azimuths (marked by '+' of given color). But eye gazes of all audio-visual training trials go after 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.

Some results (*Obr. 2:*) suggest that reference frame of ventriloquism aftereffect is mixed of fixated to eye gaze and head azimuth (E), but some results suggest that this reference frame is purely fixated to head rotation (F). In order to resolve this inconsistency, we attempted to model experimental data (*Obr. 3:*).



Obr. 3: (This figure continues.) 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 *Obr. 2:* A, B, E, F, except of yellow lines only because of technical errors and outliers removal.



Obr. 4: Continuation of Obr. 3:.

3 Newly observed phenomenon

While inspecting experimental data underlying early study in specific visualization (Figure displaying no-shift data for both center and periphery) we observed inconsistency (different one from aftereffect magnitude figure) between experimental cases, and this phenomenon is new.

We can see in (figure of no shift bias) that there are two types of such 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.

In order to explain this 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. So called training region is the area where the auditory components of audiovisual trials are within.).

4 Modelling

4.1 Model 1

In this section there is the attempt to model newly observed phenomenon, and also test of relevant qualities of its result.

4.1.1 Description

Basic idea of this modelling is to consider the concept of saccade hypometry in mathematically additive relation with concept of outside training region expansion. Saccade hypometry moves the response from position of hypothetic correct localization towards fixation point and the absolute hypometric bias increases with absolute distance from fixation point. Outside training region expansion has zero value inside training region and its absolute value increases with distance from this region.

Established variables and functions:

t ...target azimuth

$b(t, FP, trreg)$...bias of response from auditory stimulus location to target at azimuth t from eyes initially fixated at FP when the training region was $trreg$ (predicted variable).

$eotr(t, trreg)$...expansion of response on target with azimuth t outside training region $trreg$

$h(t, FP)$...saccade hypometry on target t with eyes initially fixated on FP

$dtr(t, trreg)$...relative distance of target t from training region $trreg$

Free parameters: ak_1, ak_2, ek_1, ek_2

Established equations:

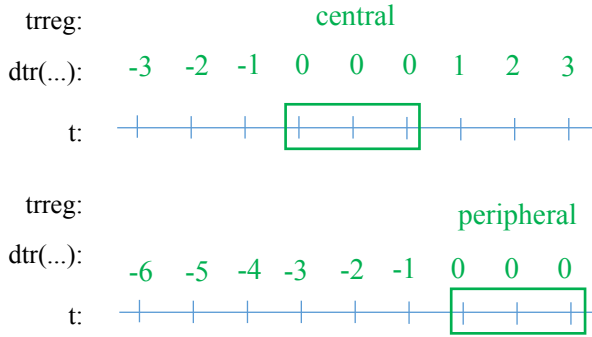
$b(t, FP, trreg) = eotr(t, trreg) + h(t, FP)$; ... additive relation of given concepts.

$eotr(t, trreg) =$

$$= ek_1 \cdot \left(\frac{1}{1 + e^{ek_2 \cdot (dtr(t, trreg))}} - \frac{1}{2} \right);$$

$h(t, FP) =$

$$= ak_1 \cdot \left(\frac{1}{1 + e^{ak_2 \cdot (t - FP)}} - \frac{1}{2} \right);$$



Obr. 5: Illustration of the relative distance of t from training region $trreg$: $trreg$ is training region, t is target azimuth and dtr is relative distance of t from training region $trreg$. (There is arithmetical distinction between this and the addend of the model, but that depend on this, according to equation in current subchapter)

4.1.2 Performance

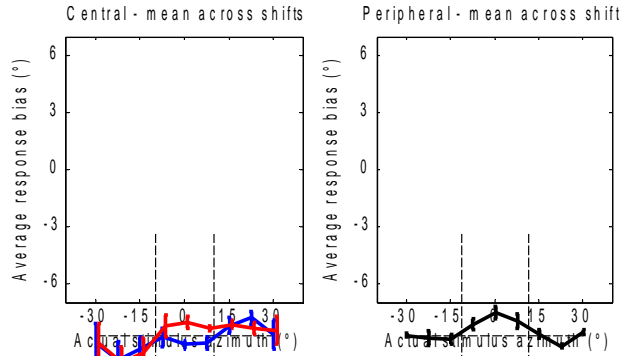
Resulting coefficients:

$$ak_1 = 0.94,$$

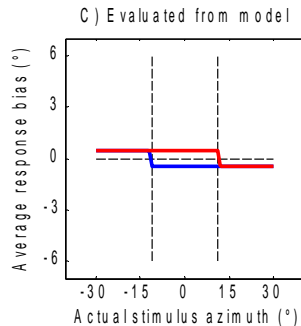
$$ak_2 = 151.13$$

$$ek_1 = -2.26$$

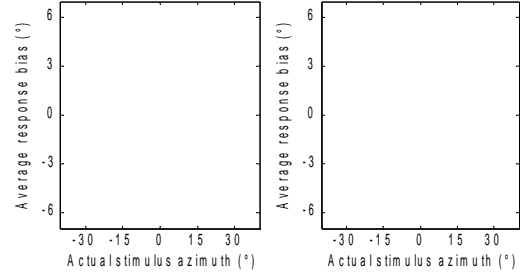
$$ek_2 = 130.62$$



Obr. 6: The visualization of the expansion outside training region according ($eotr$) to the resultant model fitted on experimental data.



Obr. 7: The visualization of the saccade hypometry ($eotr$) according to the resultant model fitted on experimental data. Colors have meaning analogical to whole article.



Obr. 8: Approximation of behavior in no-shift condition for central (left graph) and peripheral (right graph) adaptation according to given model. Colors have meaning analogical to whole article.

Figure of current modelling results (Obr. 8:) show no difference between FP dependences of central vs. peripheral adaptation (FP dependence is difference between training vs. non-training fixation biases. (subtraction of blue line from red one)). Absence of such difference is inconsistent with experimental data (Obr. 2: 1st row (no-shift)).

4.1.3 Proof of current model inappropriateness

But we can even prove mathematically that current model cannot explain current newly observed phenomenon.

We define:

$FPdep(t, trreg)$...FP dependence of biases on training region $trreg$

$TrFP$...azimuth of training fixation point.

$NtrFP$...azimuth of non-training fixation point

$$FPdep(t, trreg) =$$

$$= b(t, TrFP, trreg) - b(t, NtrFP, trreg);$$

We infer:

$$FPdep(t, trreg) =$$

$$= b(t, TrFP, trreg) - b(t, NtrFP, trreg) =$$

$$= eotr(t, trreg) + h(t, TrFP) -$$

$$(eotr(t, trreg) + h(t, NtrFP)) =$$

$$h(t, TrFP) - h(t, NtrFP);$$

We see that according to current model FP dependence ($FPdep$) does not depend on training region ($trreg$), so the object of the proof holds.

5 Conclusion

We have shown early study discussing the reference frame of the ventriloquism aftereffect and we described

partial results, which content some ambiguity. We looked to the part of the experimental data underlying so called early study, and we observed the new phenomenon. We made the attempt to model these data and we have proven that modelling described in the current article is inappropriate for the explanation of the newly observed phenomenon in manner of not explaining the newly observed phenomenon. The continuation of this research is still in progress.

Aknowledgement

Tento príspevok vznikol za podpory grantovej agentúry XXX v rámci grantovej úlohy YYY. (10 bodové písmo, Times New Roman).

Literatúra

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

Riadenie simulovaného robota iCubSim pomocou architektúry Agent-Space

Andrej Lúčny - Matúš Kopernický

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
Mlynská dolina, 842 48 Bratislava

lucny@fmph.uniba.sk, matus.kopernicky@microstep-mis.com

Abstrakt

K humanoidnému robotu iCub je voľne dostupný pomerne kvalitný simulátor iCubSim. Podarilo sa nám do neho implementovať riadenie na báze architektúry Agent-Space. Táto architektúra bola navrhnutá dávnejšie ako jedna z možných implementácií Minského predstavy o mysli ako spoločenstve agentov. Z technického pohľadu vychádza z Brooksovej subsumpčnej architektúry, ale spôsob komunikácie modulov systému je realizovaný na spôsob multi-agentového systému so stigmergickou komunikáciou. Funkčnosť riadenia je demonštrovaná mimo iné na príklade ovládania hlavy a očí robota sledujúceho pohyblivú loptičku na stole, ktorý je výsledkom interakcie ôsmich pomerne jednoduchých agentov.

1 Úvod

V roku 2006 založená iniciatíva EU Cognition si stanovila za cieľ združovať výskumníkov v oblasti kognitívnej vedy a vybaviť ich technickými prostriedkami, ktoré by umožnili vzájomnú spoluprácu a hlavne nezávislé overovanie výsledkov. Najvýznamnejším činom tejto iniciatívy bol integrovaný projekt RoboCub v rámci ktorého bol vyvinutý humanoidný robot iCub, ktorým bolo vybavených 14 výskumných pracovísk v EU (a ďalšie postupne pribúdajú). V minulosti boli totiž spravidla výsledky dosiahnuté na nejakom humanoidnom robotovi zložitejšej konštrukcie neoveriteľné, lebo nik iný nedisponoval rovnakým hardvérom. Nasadenie iCub-u túto situáciu zmenilo. Aby bol dosah tejto akcie ešte väčší, bol taktiež vyvinutý open-source simulátor iCubSim, ktorý je - napriek chybám, ktoré má - celkom dobrou aproximáciou reálneho sveta. Myšlienka za tým je tá, že pokiaľ niekto urobí zaujímavý výsledok v simulátore, môže sa kontaktovať s niektorým vlastníkom iCubu a overiť svoje pokusy v reáli.

So simulátorom iCubSim bol uskutočnený celý rad zaujímavých experimentov, viď www.icub.org. Časť z nich smerovala aj k výskumu riadiacich architektúr a to zväčša na báze neurónových sietí, napríklad (Peniak 2013).

Simulátor sa používa tak, že sa naprogramuje aplikácia, ktorá cez TCP ovláda jednotlivé senzory a

aktuátory simulovaného iCubu. Tento simulátor sme teda doplnili o aplikáciu, ktorá implementuje našu riadiacu architektúru Agent-Space (Lucny 2004). To znamená, že používateľ môže do nej vložiť svoju sadu agentov, z interakcie ktorých riadenie povstáva v zmysle (Minsky 1986), (Books 1999), (Kelemen 2003). To si žiada taktiež prekryť senzory a aktuátory ovládaním, ktoré je kompatibilné s touto architektúrou.

Túto implementáciu sme potom overili na jednoduchšej úlohe sledovania objektu podľa farby. Táto úloha však nie je triviálna natoľko, aby pri nej nevynikli jednotlivé aspekty riadiacej architektúry.

2 Simulátor iCubSim

iCubSim bol vyvinutý v C++. Dá sa inštalovať aj na MS Windows, ale jeho primárna platforma je Linux. My sme použili Ubuntu, kde sa nainštaluje veľmi ľahko:

```
# sudo sh -c 'echo "deb http://www.icub.org/ubuntu trusty contrib/science" > /etc/apt/sources.list.d/icub.list'
# sudo apt-get update
# sudo apt-get install icub
```

(trusty tu znamená len verziu Ubuntu). Okrem toho je treba v `/usr/share/icub ... context/simconfig/iCubPartsActivation.ini` nastaviť `objects = on` a pre bežný počítač `ode_param - timestamp` na cca 35 (default 10 vyžaduje výkonný hardware).

Z ďalších prostriedkov potrebujeme kompilátor GNU GCC, cmake a nejaký editor pre C++ alebo IDE ako QtCreator.

Jadrom simulátora je rovnaká softvérová platforma akú používa iCub - Yet Another Robotics Platform (YARP). V prípade simulátora je však miesto senzorov a aktuátorov robota prepojená na analogickú virtuálnu realitu vo fyzikálnom engine Open Dynamics Engine (ODE).

Simulátor iCubSim implementuje virtuálneho robota, ktorý má rovnaké parametre ako robot iCub: je 105cm vysoký, váži 20,3kg a má 53 DoF (degree of freedom), t.j. stupňov voľnosti, ktoré zahŕňajú 12 kontrolovateľných DoF na nohách, 3 kontrolovateľné DoF na torze, 32 DoF pre ruky a 6 DoF pre hlavu. Robot taktiež obsahuje rôzne druhy senzorov ako

kamery v očiach, mikrofóny a tlakové senzory (zodpovedajúce hmatu po celom tele).

Prepojenie senzorov a aktuátorov s riadiacou aplikáciou je realizované cez TCP server yarpserver, ktorý implementuje komunikačný protokol YARP. Tento server len sprostredkúva komunikáciu medzi jednotlivými klientami, pričom jedným z nich je simulátor virtuálnej reality na báze OpenGL a ODE s modelom robota iCub a prípadne zopár ďalšími predmetmi (stôl, loptička, ..., atď). yarpserver pracuje na TCP porte 10000 a umožňuje klientským aplikáciám nadviazať spojenie cez niekoľko ďalších portov tesne nad 10000 a komunikovať vzájomne YARP protokolom. Ten si možno predstaviť ako ďalšiu transportnú vrstvu umožňujúcu dátovú výmenu medzi jednotlivými modulmi klientských aplikácií (v prvom rade medzi riadiacou aplikáciou a jednotlivými senzormi a aktuátormi simulovaného robota). Je to ako keby sme nad TCP vybudovali ešte jeden protokol podobný UDP, v rámci ktorého zavádzame vlastné porty, tzv. YARP porty, akurát ich identifikátorom nie je číslo, ale text podobný ceste k súborom na disku - napríklad kamere ľavého oka je priradený YARP port /icubSim/cam/left. Yarpserver sa spustí napríklad:

```
# gnome-terminal -e yarpserver
```

Samotný simulátor sa potom spustí ako ďalší proces:

```
# gnome-terminal -e iCub_SIM
```

YARP protokol umožňuje prepojiť tok dát medzi dvomi YARP portami. Takže keď spustíme aplikáciu na zobrazovanie obrazu a priradíme jej nami definovaný port /camera:

```
# gnome-terminal -e "yarpview /camera"
```

neukazuje nič, kým do nej nepresmerujeme dáta z kamery v yarpserveri, ktorá beží na porte /icubSim/cam/left, ktorý je definovaný v simulátore iCubSim:

```
# yarp connect /icubSim/cam/left /camera
```

Na každý YARP port sa pritom môže pripojiť viacero aplikácií, takže tým, že si obraz z ľavého oka zobrazíme v yarpview, nestrácame možnosť brať jeho hodnotu rovnakým spôsobom do riadiacej aplikácie.

Architektúra riadiacej aplikácie je ponechaná na jej vývojára, ktorý zo svojho programu volá funkcie klientskej knižnice implementujúcej YARP protokol. Akú architektúru zvolí, je teda len na ňom – spravidla píše jednovláknový program synchronne volajúci spomínanú klientskú knižnicu. Obraz z kamery ľavého oka získame napríklad takto:

```
localPort = "/left/image/in";  
imagePort = new yarp::os::BufferedPort  
<yarp::sig::ImageOf<yarp::sig::PixelRgb>>();
```

```
imagePort->open(localPort);  
  
remotePort = "/icubSim/cam/left";  
yarp::os::Network::connect(remotePort,localPort);  
  
yarp::sig::ImageOf<yarp::sig::PixelRgb> *image;  
image = imagePort->read();  
if (image != NULL) process(image);
```

Podobne pošleme príkaz na motor:

```
yarp::dev::PolyDriver driver;  
yarp::dev::IPositionControl *pos;  
yarp::dev::IVelocityControl *vel;  
yarp::dev::IEncoders *encs;  
yarp::dev::IControlLimits *iLim;  
yarp::sig::Vector speeds;  
yarp::sig::Vector accelerations;  
string Device = "remote_controlboard";  
string LocalPort = "/test/client";  
string RemotePort = "/icubSim/head";  
yarp::os::Property options;  
options.put("device", Device);  
options.put("local", LocalPort);  
options.put("remote", RemotePort);  
driver.open(options);  
driver.view(encs);  
driver.view(pos);  
driver.view(iLim);  
// compute speeds & accelerations  
pos->setRefAccelerations(accelerations.data());  
pos->setRefSpeeds(speeds.data());
```

Keď implementujeme nejaké konkrétne správanie, musíme napísať takú aplikáciu, ktorá zo sekvencie vstupov zo senzorov a propriocepce počíta v reálnom čase adekvátne výstupy na aktuátory.

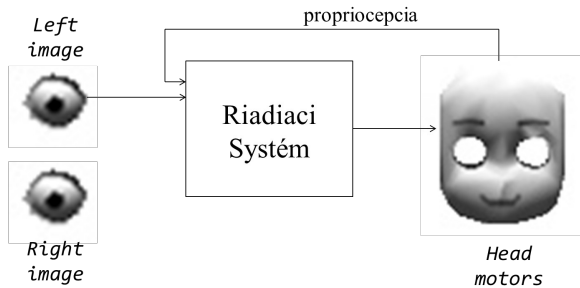
3 Modelová úloha

Za modelovú úlohu sme si zvolili sledovanie loptičky danej farby, upriamenie pozornosti na ňu a jej udržanie v zornom poli robota (obr. 1).



Obr. 1: Virtuálna scéna s iCubom a objektmi

Keď túto úlohu premietneme do vzťahu senzorov a aktuátorov, ide o spracovanie obrazu z oka (berieme do úvahy len jedno) na pohyb štyroch motorov: dva hýbu hlavou zľava doprava a zhora dole a dva hýbu očami zľava doprava a zhora dole (obr. 2).



Obr. 2: Riadiaci systém vo vzťahu k senzorom a aktuátorom.

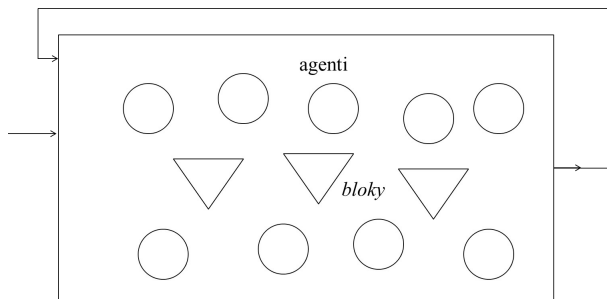
4 Riadiaca architektúra Agent-Space

Pokiaľ by sme sa modelovú úlohu snažili vyriešiť tradičným spôsobom, v jednom vlákne by sme nejakým cyklickým hýbali motory, až kým by sa loptička nedostala do správneho zorného poľa. Približný kód takého programu vidíme na obr. 3.

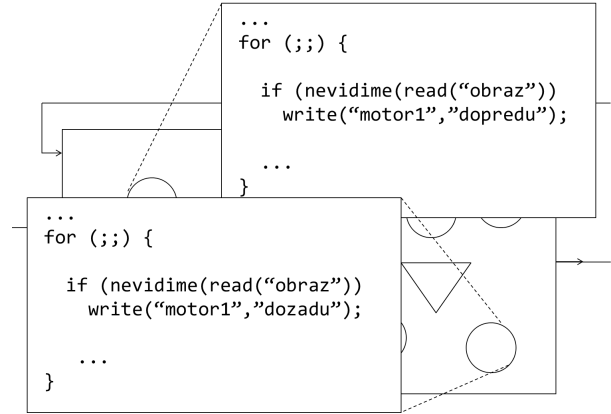
```

for (;;) {
  for (;;) {
    for (;;) {
      for (;;) {
        if (lopticka(obraz()) break;
        else pohni(motor4);
      }
      if (lopticka(obraz()) break;
      else pohni(motor3);
    }
    if (lopticka(obraz()) break;
    else pohni(motor2);
  }
  if (lopticka(obraz()) break;
  else pohni(motor1);
}
  
```

Obr. 3: Tradičný riadiaci systém



Obr. 4: Elementy riadenia podľa architektúry Agent-Space



Obr. 5: Povstávanie riadenia z viacerých jednoduchších správani

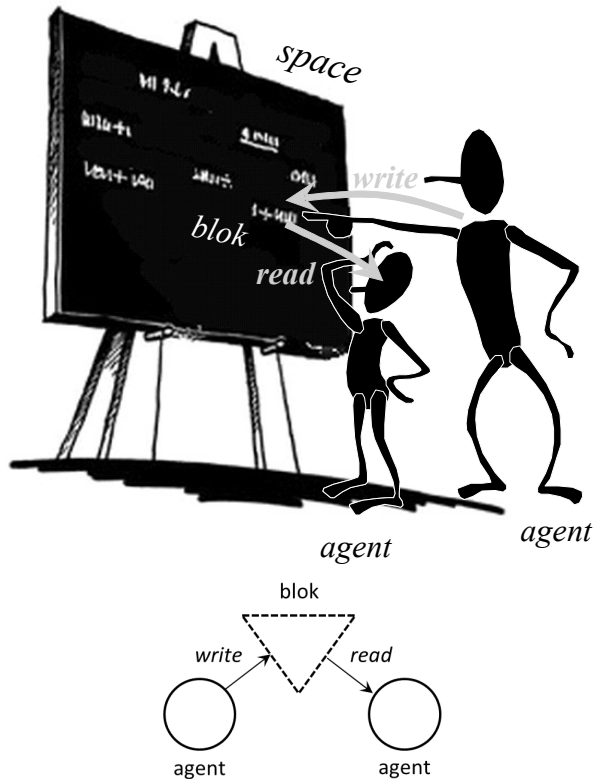
Aby pohyb hlavy a očí nebol príliš mechanický a drevorubačský, prispôbovali by sme hýbanie motoru vnútornému stavu systému, ktorý by sme premietli do hodnôt globálnych premenných.

Inovácia nášho prístupu spočíva v tom, že riadiaci systém bude mať viac vlákien, medzi ktorými bude prebiehať interakcia, čiže normalizovaná dátová výmena. Každé z týchto vlákien budeme považovať za samostatne konajúceho agenta, schopného interakcie s ostatnými agentmi prostredníctvom tzv. blokov (stigiem) na čiernej tabuli, či nástenke. Implementujeme tak societu agentov, medzi ktorými prebieha tzv. stigmergická komunikácia (obr. 4). Pritom každý jeden agent je z hľadiska riadenia porovnateľný s celým riadiacim programom pri tradičnom riešení, akurát, že jeho kód by mal byť podstatne jednoduchší – čo je zmyslom takejto dekompozície riadenia. Programy agentov sa nemusia vždy len dopĺňať, môžu taktiež súperiť, t.j. pokúšať sa uskutočniť aj úplne protichodné akcie na aktuátoroch (obr. 5).

Každý z týchto agentov má vlastné riadenie nezávislé na ostatných agentoch a vykonáva vlastný kód, pričom tento beží v nekonečnom cykle. Pri každom prechode týmto cyklom (ktorý vyvoláva určitá udalosť a to buď impulz od časovača alebo dostupnosť nových informácií) sa na základe vnímania aktuálnej situácie a ďalších informácií uložených v tzv. vnútornom stave agenta, vypočíta a vykoná náležitá akcia.

Najvýznamnejšou zložkou vnímania je pre každého agenta komunikácia s inými agentmi. Práve tento komunikačný mechanizmus zaraďuje agenta do systému. V použitej architektúre Agent-Space je tento mechanizmus založený výlučne na nepriamej komunikácii a je realizovaný tzv. čiernou tabuľou (nástenkou), ktorú nazývame space (obr. 6). Prostredie dokáže uskladňovať pomenované dáta, tzv. bloky, ktoré agenti dokážu čítať, zapisovať a mazať. Agenti pritom musia poznať ich mená a formát dát v nich zapísaných.

Detaily tejto architektúry, ktoré je potrebné zväziť, aby tento koncept fungoval naozaj dobre, sú popísané v (Lucny 2004) (Lucny 2012). Tu spomenieme len základnú povahu zápisu blokov, pomocou ktorej sa dá realizovať dátový tok od mnohých producentov ku mnohým konzumentom.



Obr. 6: Komunikácia medzi agentmi cez space a jej symbolické vyobrazenie

Bloky sa v space vytvárajú prvým zápisom. Čítať ich však možno už predtým a to bez toho, že to vyvolalo nejakú výnimku: pre tento prípad je agent pri čítaní povinný zdefinovať preddefinovanú hodnotu, ktorá sa pri čítaní neexistujúceho alebo prázdneho bloku vráti, ako keby v ňom bola zapísaná.

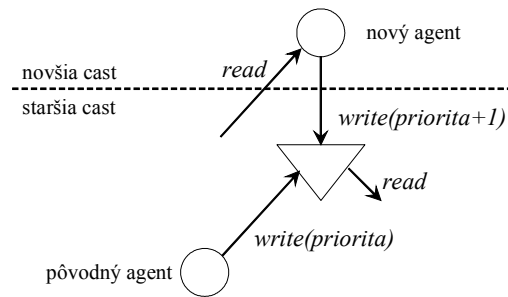
Blok môže obsahovať iba jedinú hodnotu, takže hodnota zapísaná jednou operáciou zápisu je prepísaná nasledovnou. Hodnota zapísaná v bloku je prepísaná bez ohľadu na to, či ju nejaký agent stihol prečítať alebo nie. Takže ak producent zapíše nasledujúcu hodnotu do bloku prv, ako niektorý konzument stihne blok prečítať, tak tento konzument o túto hodnotu proste príde – z jeho pohľadu nikdy v bloku nebola. Takže ak jeden konzument nie je dostatočne rýchly, aby stihol z bloku prevziať všetky zapisované hodnoty, automaticky tu príde k ich vzorkovaniu, zatiaľ čo druhý ich môže stihnúť prečítať všetky. Vďaka tejto vlastnosti je v systéme ľahké kombinovať pomalé a rýchle moduly.

Pri každom zápise je možné taktiež definovať časovú platnosť zapísanej hodnoty, t.j. dobu, po uplynutí ktorej hodnota z bloku automaticky zmizne

(čo zariadi space). Pokiaľ časová platnosť pri zápise nie je definovaná, hodnota ostáva v bloku natrvalo, respektíve do najbližšieho prepísania. Časová platnosť je dobrý nástroj na vyjadrenie dočasného charakteru informácie. Navyše podporuje dátový tok od mnohých producentov k mnohým konzumentom. Keď máme napríklad viac metód na výpočet určitej hodnoty, pričom každá vyprodukuje hodnotu len za určitých podmienok, môžeme tieto metódy premeniť na agentov zapisujúcich rovnaký blok. Takto sa konzumenti nemusia starať z akého zdroja hodnota pochádza. Musíme však zariadiť, aby agent, ktorý nič nespočítal, radšej nič do bloku nezapísal, lebo tak by prepísal použiteľnú hodnotu od iného agenta. Potom ale zákonite hrozí, že nik nespočíta nič a v bloku naveky ostatne stará posledná hodnota. Preto v takomto prípade treba zdefinovať jej časovú platnosť.

Okrem časovej platnosti hodnoty, je možné pri zápise hodnoty zdefinovať aj jej prioritu. Toto nám umožňuje preferovať hodnotu od určitého producenta. Priorita zápisu je realizovaná v space tak, že keď sa nejaký agent snaží prepísať v space prioritnejšiu hodnotu, hodnota sa nezmení (pričom on o tom nie je nijako informovaný). Tento mechanizmus je dôležitý, keď sa pri inkrementálnom vývoji snažíme z novo pridaných častí systému ovplyvniť pôvodné (obr. 7), čo je podstata Brooksovej subsumpcnej architektúry (Brooks 1999). Pri agent-space sa novo pridaní agenti miešajú do komunikácie medzi pôvodnými tak, že:

- noví čítajú bloky, cez ktoré si pôvodní vymieňajú informácie (odpočúvanie)
- noví prepisujú hodnotu v týchto blokoch (supresia)
- noví vymazávajú hodnotu v týchto blokoch (inhibícia)



Obr. 7: Supresia starších častí z novších

Vieme teda vyjadriť všetky mechanizmy Brooksovej subsumpcie (odpočúvanie, supresiu a inhibíciu) a dokonca aj komplikovanejšiu koordináciu nových a pôvodných agentov. To nám umožní organizovať komplexnejší systém, ktorého zložité správanie necháme vzniknúť z interakcie jeho jednoduchých častí, ktorých správanie vieme zdefinovať ľahšie.

5 Implementácia Agent-Space do iCubSim

Na implementáciu architektúry Agent-Space do iCubSim je okrem implementácie všeobecného agenta (implementuje sa ako objekt s vlastným vláknom (kód 1)) a mechanizmu komunikácie medzi agentmi (implementuje sa cez statickú pamäť pomocou synchronizácie vlákien, viď (Lucny 2012)) treba pre každý senzor (vrátane propriocepcie) a aktuátor implementovať blok, pomocou ktorého ho je možné monitorovať či riadiť.

```
#include "agentspace.h"

class MyAgent : public Agent {
protected:

    void init (string args) {
        //timer_attach(period,period);
        //id=trigger_attach("name");
    }

    void sense_select_act (int id) {
        //value = space_read("name",default);
        //space_write("name",value);
        //space_write("name",value,time_validity);
        //space_write("name",value,time_validity,priority);
    }

public:
    MyAgent (string args) :
        Agent(args) {};
};
```

Kód 1. Všeobecný kód implementujúci agenta

Pritom sme sa zamerali na podporu takej povahy ovládania senzorov, aby zodpovedala potrebám práce v reálnom čase. Bežne sa totiž pri práci s iCubom i jeho simulátorom ovládajú aktuátory príkazmi typu "zmeň tento kĺb do tejto polohy", zatiaľ čo našim zámerom vyhovuje skôr derivačný pokyn "zmeň smer a rýchlosť pohybu tohto kĺbu takto". Miesto "pohnúť" používame "hýbať" - miesto jedného pokynu aktuátoru, sériu pokynov rozloženú v čase, miesto invocácie, reguláciu. iCubSim nám tu našťastie výrazne pomáha tým, že fyzické limity pohybu kĺbov stráži sám - riadiaci systém si teda môže želať akýkoľvek pohyb, ale ak už je daný kĺb v krajnej polohe, tak sa pohyb neudeje. Vydanie pokynu pre nás pritom znamená zapísanie určitej hodnoty rýchlosti alebo zrýchlenia do určitého bloku v space. Čítaním iného bloku si môžeme overiť skutočnú aktuálnu polohu kĺbu.

Práca so senzormi je o niečo jednoduchšia, lebo stačí zariadiť, aby určitý blok v space obsahoval poslednú hodnotu zo senzora, odkiaľ si ho môžu agenti kedykoľvek prečítať.

Funkčnosť bloku v space, ktorý reprezentuje senzor či aktuátor je najľahšie zabezpečiť agentom, ktorý bude

potom súčasťou každej riadiacej aplikácie, ktorá daný senzor či aktuátor používa. Pre našu modelovú úlohu sme potrebovali jednak obraz z kamery oka, čo riešil CamSensAgent (kód 2), jednak ovládanie motorov očí a krku, čo riešil VelocityMotorAgent (kód. 9).

```
void CamSensorAgent::init(string args) {
    remotePort = "/icubSim/cam/left";
    localPort = "/left/image/in";
    blockName = "leftImage";
    startTime = 1000; period = 250;
    imagePort = new yarp::os::BufferedPort
        <yarp::sig::ImageOf<yarp::sig::PixelRgb>>();
    imagePort->open(localPort);
    yarp::os::Network::connect(remotePort,localPort);
    timer_attach(startTime,period);
}

void CamSensorAgent::sense_select_act(int id) {
    yarp::sig::ImageOf<yarp::sig::PixelRgb> *image
        = imagePort->read();
    if (image!=NULL) space_write(blockName, image);
    else printf("No image\n");
}
```

Kód2. Príjem obrazu z kamery do bloku v space

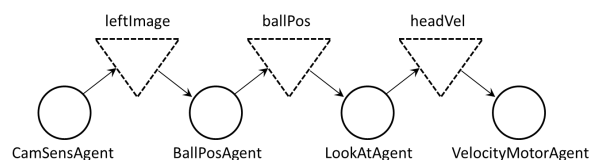
```
void VelocityMotorAgent::init(string args) {
    ...
    string velocityBlock = "head_vel";
    yarp::dev::PolyDriver vel;
    yarp::os::Property options;
    options.put("device", deviceName);
    options.put("local", LocalPort);
    options.put("remote", remotePort);
    vel.open();
    ...
    trigger_attach(velocityBlock);
}

void VelocityMotorAgent::sense_select_act (int id) {
    yarp::sig::Vector commands;
    commands = space_read(velocityBlock, def_commands);
    vel->velocityMove(commands.data());
}
```

Kód 3. Ovládanie motorov podľa bloku v space

6 Implementácia modelovej úlohy

Implementáciu architektúry sme potom otestovali na modelovej úlohe. Pri vývoji jej riešenia sme postupovali v zmysle inkrementálneho vývoja zdola nahor, typického pre Brooksovu sumbumpčnú architektúru (Brooks 1999).



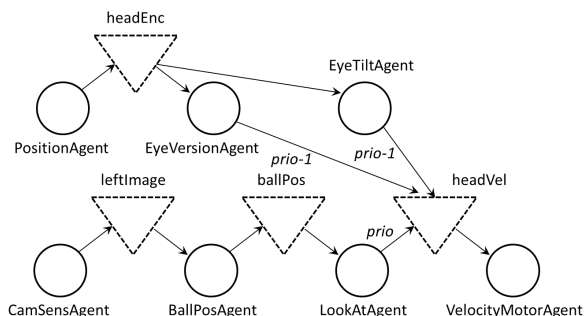
Obr. 8: Vývojová fáza 1

Prvá fáza vývoja rieši základnú situáciu, že robot sledovanú loptičku vidí, ale nemá ju v strede zorného poľa (obr.8). V takom prípade z pozície loptičky na obraze vypočítame vektor rýchlosti. Tu si treba uvedomiť, že vektor rýchlosti sa vypočíta oveľa ľahšie než výsledná poloha. Tá sa nájde vďaka spätnej väzbe ako poloha v okamihu, kedy je loptička v strede obrazu. Tento prístup môžeme interpretovať ako tzv. stelesnenie (embodiment) - telo robota sa tu podieľa na samotnom výpočte správnej polohy. Riadenie zabezpečujú dvaja agenti: BallPosAgent detekuje polohu loptičky na obraze (kód 4) a LookAtAgent podľa danej polohy voči stredu obrazu určuje smer rýchlosti.

```
void BallPosAgent::init(string args) {
    inBlock = "leftImage";
    outBlock = "ballPos";
    trigger_attach(inBlock, TRIGGER_MATCHING);
}
```

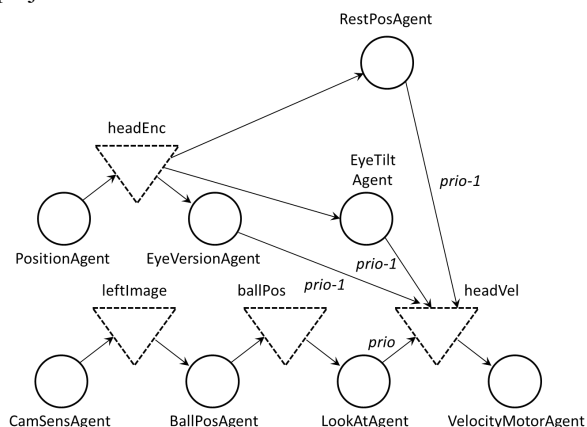
```
void BallPosAgent::sense_select_act(int pid) {
    int xMean=0, yMean=0;
    yarp::sig::ImageOf<yarp::sig::PixelRgb> *image;
    image = space_read(conf.inBlock, defimage);
    if (image != NULL) {
        width = image->width();
        height = image->height();
        for (int x = 0; x < width; x++) {
            for (int y = 0; y < height; y++) {
                yarp::sig::PixelRgb& pixel;
                pixel = image->pixel(x,y);
                if (pixel.b > pixel.r*1.2+10
                    && pixel.b > pixel.g*1.2+10) {
                    xMean += x; yMean += y; count++;
                }
            }
        }
    }
    if (count > 0) {
        xMean /= count; yMean /= count;
        target.x = xMean; target.y = yMean;
        space_write(outBlock, target, 300);
    }
}
```

Kód 4. Agent detekujúci polohu loptičky na obraze



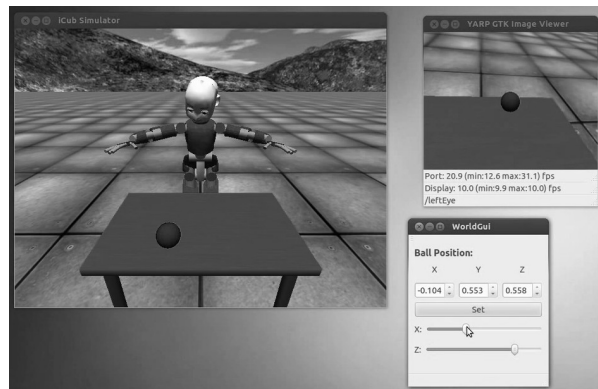
Obr. 9: Vývojová fáza 2

Druhá fáza (obr. 9) rieši situáciu, keď robot loptičku v zornom poli nemá. Vtedy začína gúľať očami, pričom miesto nejakého globálneho algoritmu prehľadávania, používa jednoduchý trik a to, že nezávisle na sebe nimi pohybuje zľava doprava (EyeVersionAgent) a zhora dole (EyeTiltAgent) rôznymi rýchlosťami a smer pohybu mení na základe dosiahnutia krajných polôh, o čom sa dozvie z propriocepcie, ktorú zabezpečuje PositionAgent. Následkom toho je pohyb očí oveľa prirodzenejší, než keby sa robot snažil organizovane prehľadať priestor okolo seba. Pritom sa netreba báť, že by sa loptička ocitla v nejakom mŕtvom priestore, ktorý robot do pohľadu nedostane, hoci by na ňu dovidel. Pokyny od agentov gúľajúcich očami (EyeVersionAgent, EyeTiltAgent) majú menšiu prioritu ako pokyny od sledovania loptičky (LookAtAgent), takže pokiaľ robot loptičku vidí, tak očami negúľa. Na druhej strane tu vidíme, že je dôležité, aby mali pokyny od LookAtAgent obmedzenú časovú platnosť, aby sa gúľanie očí vôbec niekedy prejavilo.



Obr. 10: Vývojová fáza 3

V tretej fáze sme pridali pohyb hlavy, ktorý je potrebný, keď gúľanie očami nestačí. Ten realizuje RestPosAgent, ktorý hýbe hlavou zľava doprava (obr. 10).



Obr. 11: Implementácia motivačnej úlohy

7 Záver

V príspevku sme prezentovali rozšírenie simulátora iCubSim o možnosť tvoriť jeho riadiacu aplikáciu na báze architektúry Agent-Space. Vo vyvinutom prostriedku sme potom implementovali ako príklad sledovanie loptičky robotom, pričom užívateľ mohol touto loptičkou hýbať, aj ju schovať (obr. 11). Zdrojové kódy ako aj videá prezentujúce výsledok sú na www.agentspace.org/icub.

V budúcnosti sa budeme snažiť jednak trochu tieto kódy očistiť a zdokumentovať, jednať previazať vnímanie robota s externým prostredím, aby videl nielen predmety na svojom stole, ale aj samotného experimentátora, t.j. užívateľa programu.

Literatúra

- R. A. Brooks: *Cambrian Intelligence*, The MIT Press, Cambridge, Mass., 1999.
- Kelemen, J.: *The Agent Paradigm. Computing and Informatics*, Vol.22. (2003), pp. 513-519
- Lúčny, A.: *Building Complex Systems with Agent-Space Architecture. Computing and Informatics*, Vol. 23 (2004), pp. 1001-1036
- Lúčny, A.: *Od medzimodulových spojení k nepriamej komunikácii medzi agentami. Znalosti*, VŠE Ostrava, 2007.
- Lúčny, A.: *Multiagentový prístup k modelovaniu mysle - alebo ako sledovať pingpongovú loptičku*. In: *Umelá inteligencia a kognitívna veda III. (Kvasnička V. ed.)*, STU, Bratisva, 2011
- Lúčny, A.: *Otvorená implementácia architektúry Agent-Space*. In: *Kognice a umělý život XII. (Kvasnička, V. - Wiedermann, J. eds.)*, Agentura Action M, Praha, 2012, pp. 132-136
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C., Rosander, K., Lopes, M., Santos-Victor, J., Bernardino, A. a Montesano, L. (2010). *The icub humanoid robot: An open-systems platform for research in cognitive development. Neural Networks*. vol. 23, no. 8-9, pp. 1125–1134.
- Minsky, M.: *Society of Mind*. Simon & Schuster, New York, 1986
- Peniak, M. (2013). *Aquila 2.0 Software Architecture for Cognitive Robotics. ICDL 2013*

Implementace a synchronizace Hayekova stroje pro řízení robota Khepera IV

Lukáš Mamula

Slezská univerzita v Opavě, Filozoficko-přírodovědecká fakulta v Opavě
Bezručovo náměstí 13, 74601 Opava
mamula.lukas@gmail.com

Abstrakt

Díky použitému robotu Khepera IV si můžeme v tomto článku ukázat, jak implementovat kooperační mechanismus tak, aby bylo možno ovládat robota bez lidského zásahu. Článek vychází z předchozích teoretických studií, které jsme museli provést a které nám pomohli pochopit fungování multiagentových systémů. V článku se velice okrajově zaměříme na teoretický popis našeho multiagentového systému a spíše se zaměříme na technický popis proveditelnosti implementace. Z pozorování chování robota, který je řízen pomocí Hayekova stroje, můžeme vyčíst velice zajímavé stavy. Právě řešení stavů, do kterých se náš robot dostane, bude klíčové při našem bádání. Zda některé kognitivní funkce přisuzované živým bytostem, mohou vznikat i v rozsáhlých systémech, jako určitý druh emergence. Je možné, aby robot řídící se softwarem, jehož princip je v našem případě založen na ekonomických mechanismech, mohl nějakým elementárním způsobem vykazovat známky kognitivního chování? Věříme, že na tuto otázku dostaneme odpověď v průběhu našeho snažení a v tomto článku se této otázce budeme také věnovat.

1 Úvod

Umělá inteligence a přístupy k této problematice spadají do vědní disciplíny, která si zaslouží obzvlášť velkou pozornost. I v tomto článku se podíváme na jeden z možných způsobů, jak dosáhnout takového jevu. Jevu, který se pozorovateli fyzického zařízení (robota) nebo softwarové aplikace, může jevit jako samostatně fungující entita.

V našem článku se zaměříme na vývoj umělé inteligence pomocí konceptu Hayekova stroje. Tento stroj je značně odlišný od toho, co popisuje ve své knize A. Kubík (2004). Náš koncept vychází z diplomové práce, která je také uvedena v literatuře. V

první kapitole se seznámíme s krátkým popisem kooperačního mechanismu, kterým se snažíme dosáhnout inteligentního chování. Ať už při manipulaci s robotem nebo v simulátoru.

Po seznámení čtenáře se základy naší práce, se můžeme zaměřit na pozorování chování. Chování systému se odlišuje podle toho, zda systém spouštíme v režimu simulátoru nebo zda používáme fyzického robota Khepera IV. V obou zmiňovaných režimech využíváme stejný kooperační mechanismus, avšak získaná data k následnému zpracování se dost liší.

2 Aktuální stav

Jak samotný název této kapitoly napovídá, zaměříme se na současný stav naší aplikace. V úvodu jsme nepřímo naznačili, že aplikace je rozdělena do dvou režimů: *simulátor a reálný provoz*.

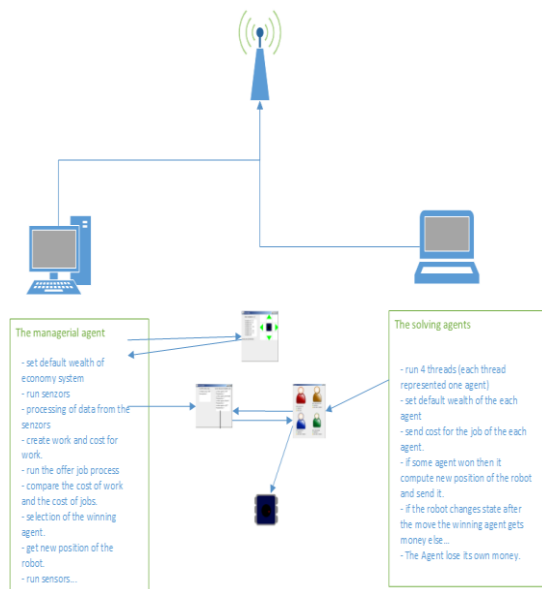
K rozdělení došlo z důvodu časových a výkonnostních. Jak již bylo řečeno v obou režimech je používán stejný kooperační mechanismus. Rozdíl spočívá ve spektru získaných informací ze simulátoru a reálného prostředí. Avšak rozdíl je také ve zpracování některých informací. Při tvorbě simulátoru jsme se snažili napodobit získávání dat tak, aby byl tento proces co nejvíce podobný sbírání dat z reálného prostředí. Z toho také vyplývá, že simulátor byl vytvářen jako první a následně upravován pro reálný provoz.

V současnosti naše aplikace umí ovládat pohyb robota prostředím, vyhýbat se překážkám a zaznamenávat prostředí do jednoduché struktury. Ačkoliv se může zdát, že toho není příliš mnoho, je potřeba si uvědomit, že tohoto pozorovaného chování, jsme dosáhli za pomoci ekonomických principů aplikovaných v kooperačním mechanismu. Tedy pohyb robota je jev, který vznikl emergencí vnitřních událostí a datových toků systému.

Nyní se přesuneme k první části této kapitoly, kde si velice jednoduše popíšeme nastavení systému a kooperační mechanismus, pomocí něhož zajišťujeme výsledný pohyb robota ať už v reálném nebo virtuálním prostředí. Podrobněji je tento mechanismus popsán *L. Mamula (2014)*.

2.1 Úvodní nastavení systému

Pro popis kooperačního mechanismu použijeme Obr. 1. Na tomto obrázku můžeme vidět jednoduchou topologii, ve které máme dva počítače. Druh počítačů pro naše účely není až tak důležitý.



Obr. 1: Popis kooperačního mechanismu

Druhým způsobem, jak může náš systém pracovat, je spuštění celého systému na jednom PC. Tato varianta je možná u počítačů s vyšším počtem jader (např. iCore 3). Vysvětlíme si, proč je to takto řešeno.

Naše aplikace je rozdělena na manažerského agenta a řešitelské agenty. Vždy když spouštíme naši aplikaci, spouštíme dva programy reprezentující právě výše zmíněné agenty.

V případě aplikace s manažerským agentem se nejedná o výpočetně náročnou aplikaci, avšak v případě aplikace s řešitelskými agenty je problém opačný.

Každý agent je reprezentován spuštěním samostatného vlákna tzv. Thread. Pokud spouštíme aplikaci s řešitelskými agenty, jsou na začátku vytvořeny čtyři vlákna, které se následně množí dle mechanismu, jenž si vysvětlíme později. Proto je potřebné zajistit

výpočetní výkon dostatečně veliký, aby nedocházelo k velkým časovým prodávám v komunikaci mezi aplikacemi tzn. mezi manažerským a řešitelskými agenty.

Pokud spustíme celý systém v prostředí jednoho počítače, není potřeba nastavovat síťové připojení a aplikace spolu komunikují skrze lokální soubor, kde si zanechávají zprávy o akcích agentů (bohatství, úspěšnost řešení atd). V druhém případě, pokud rozdělíme systém mezi dva počítače, je potřeba upřesnit síťové nastavení (ip adresa, maska atd). V dalších krocích je fungování totožné. Aplikace spolu komunikují skrze zprávy, které zanechávají v textovém souboru.

Tímto jsme si vysvětlili, jak připravit systém pro následující komunikaci s robotem nebo softbotem. V další části této práce si ukážeme, jak vybrat, zda chceme systémem ovládat simulovaného (virtuálního) softbota nebo reálného robota, v našem případě robota Khepera IV.

2.2 Režim simulátor/robot

V této kapitole budeme předpokládat, že máme systém nastaven jedním ze způsobů tak, jak jsme si popsali v předchozí kapitole. To zda nám obě aplikace poběží na jednom PC nebo na dvou zařízeních nijak neovlivní ovládání robota nebo softbota.

Nyní musíme vybrat zda systém bude pracovat s virtuálním softbotem nebo reálným robotem. Pokud budeme používat reálného robota, musíme v aplikaci nastavit síťové připojení k robotu. Dále je potřeba dalších nastavení a to přímo v robotu. Khepera IV v sobě obsahuje pouze základní verzi linuxového systému, což je pro naše účely málo.

Jako tvůrci celého systému jsme nuceni v robotu nastavit bezdrátové připojení, s čím také souvisí vytvoření domácího prostředí s jedním acces pointem v podobě např. Wi-fi routeru. Dále je zapotřebí nastavit adresářovou strukturu na straně robota pro spuštění kodu v jazyce C/C++. S tím souvisí instalace *toolchain* nástroje od výrobce robota, což nám umožní pracovat s funkcemi, jako jsou například funkce pro ovládání motorku kol robota, či získání dat ze senzorů v podobě čísel v desítkové soustavě.

Instalovaný nástroj integrujeme do naší jednoduché třídy, kterou vhodně umístíme do adresáře robota.

Třída bude mít za úkol zajistit komunikaci s naší aplikací přes lokální síť. Komunikace bude spočívat v posílání dat ze senzorů robota a následné zaslání údajů z naší aplikace o tom, kam se má robot posunout a v pozdější fázi vývoje, co má robot dělat.

Pokud bychom chtěli používat pouze simulátor, je řešení o mnoho jednodušší. Z pohledu uživatele se změní pouze to, že nebude muset zadávat IP adresu pro připojení k robotu. Uživatel pouze umístí robota do prostředí, které mu sám dopředu v editoru prostředí vytvoří. Editor prostředí je součástí naší aplikace a obsahuje širokou škálu překážek.

Nyní máme aplikaci a také robota/softbota připravené ke spuštění. V následující kapitole si krátce vystěhlíme, co se děje v systému, než se provede pozorovaný pohyb robota.

2.3 Popis kooperačního mechanismu

Přejdeme k vysvětlení kooperačního mechanismu. V dřívějších verzích tohoto systému jsme se snažili nepoužívat genetických algoritmů. Bohužel nebo bohudík jsme dospěli po delším bádání k závěru, že bez genetických algoritmů nebudeme schopni vytvořit komplexní systém.

Jako následek tohoto zjištění bylo to, že jsme se ve vývoji vrátili k původní myšlence, kterou popsal již Aleš Kubík ve své knize. Samozřejmě pro naše účely došlo k určitým modifikacím, ale základ zůstal dosti podobný.

Jako kooperační mechanismus jsme zvolili nabídku a poptávku na trhu práce. Tímto mechanismem jsme vytvořili vnitřní procesy, které nám umožnili libovolně pohybovat původně se softbotem a následně s robotem.

Na Obr. 1 máme celý mechanismus popsán a protože jsme se o tomto mechanismu zmiňovali již v dřívějších pracích, popíšeme si tento mechanismus velice zkráceně.

Kooperační mechanismus je vytvořen ve smyslu odměňování a trestání. Na začátku jsou spuštěny senzory, z kterých se vytvoří řetězec reprezentující aktuální polohu robota. Ze získaného řetězce se vytvoří nabídka práce a cena za práci. Mechanismus vytváření ceny za práci je trochu složitější a záleží na více podmínkách. Nejdůležitější podmínkou je, zda agent svou práci splní stav robota. Stav robota jsou čtyři: *prohledává, nese, sbírá a potravu*.

Spolu s výše popsanými procesy je paralelně spuštěna aplikace, jenž reprezentuje řešitelské agenty. Z předchozí kapitoly víme, že aplikace může být umístěna na stejném PC nebo na vzdáleném zařízení v rámci stejné lokální sítě. Na začátku jsou defaultně vytvořeny čtyři agenty. Množství agentů jsme vybrali záměrně.

Na začátku se robot může pohybovat pouze čtyřmi směry: nahoru, dolů, doprava, doleva. Každý z vytvořených agentů odpovídá právě za jeden směr. V této aplikaci je jednotlivým agentům přiřazeno defaultní bohatství.

Pokud máme vytvořenou cenu za práci, manažerský agent kontroluje, zda jsou přihlášeny řešitelské agenty, které budou chtít získat nabízenou práci. Řešitelské agenty z paralelně běžící aplikace se přihlásí manažerskému agentu. Spolu s přihlášením (jméno agenta) poskytnou manažerskému agentu také cenu, za kterou svou práci budou vykonávat. Tato cena je vytvářena na základě více kritérií.

Cena, za kterou budou řešitelské agenty vykonávat svou práci se odvíjí od aktuálního bohatství. V případě řešitelských agentů se bohatství mění velmi dynamicky. Pokud některý z řešitelských agentů získá práci a vykoná svou práci, pak se ostatním agentům v systému odečte -1 od jejich aktuálního bohatství a suma odečtených hodnot od všech agentů se přičte vítěznému agentu tzn. Pokud máme na začátku čtyři agenty pak, když vyhraje jeden z nich, získá ($pocet_agentu-1$) bodů ke svému bohatství.

Nyní přichází na řadu problém genetických algoritmů. Pokud se některý z agentů dostane se svým bohatstvím až na hodnotu 0, pak tento agent zaniká resp. jeho existence v systému končí. Avšak aby nám nevymřel celý náš multiagentový systém, platí zde, že agent, který dosáhne dvojnásobku svého bohatství, požádá odpovědného agenta o vytvoření svého potomka. Ten bude dědit vlastnosti svého rodiče a dalšího agenta, který je náhodně vybrán z množiny mrtvých či méně úspěšných agentů. Agent, jenž vytváří nové agenty se neúčastní aukcí a nemá přiřazené bohatství.

Tento nový agent v sobě slučuje vlastnosti a schopnosti svých rodičů. Novému agentu se připíše polovina bohatství agenta, jež požádal o vytvoření potomka. Následně se toto bohatství odečte tomuto úspěšnému agentu. Tím zachováme neměnné množství bohatství uvnitř systému. Nový agent se v dalším životním cyklu

již účastní aukcí a snaží se získat práci, aby mohl dále získávat bohatství, z kterých by mohl vytvářet další potomky. Každý nový agent odevzdává daň svému rodiči, který dal žádost k jeho vytvoření.

Nyní se přesuneme do další kapitoly, kde si popíšeme, jak pracuje proces *výběrového řízení*. Výběr vítězného agenta, který provede práci a následně je za ní ohodnocen, tak jak jsme popsali výše.

Obr.



2.4 Proces výběrového řízení

V minulé kapitole jsme si vysvětlili, jak vytváříme nové agenty a jak agenty odměňujeme. Nyní se zaměříme na to, jak pracuje dílčí mechanismus výběru vítězného agenta, který poté provede pohyb robotem.

Pokud se systém nachází na začátku, tedy krátce po spuštění je vítězný agent vybrán náhodně z přihlášených agentů. V dalších cyklech již vybíráme na základě vnitřních pravidel.

Na začátku manažerský agent získá data ze sensorů. Z těchto dat vytvoří nabídku práce. Po přihlášení jednotlivých řešitelských agentů z paralelní aplikace, začne manažerský agent prověřovat jednotlivé přihlášené. Kontroluje se zde několik kritérií, které se následně posílají do pravidlového systému, kde se data fuzziifikují a defuzziifikují. Díky pravidlového systému, jenž tvoří několik fuzzy množin odrážející stav zasílaných dat řešitelských agentů, získáme číslo. Toto číslo manažerský agent použije při výběru vítězného agenta, který získá práci a s ní i odměnu.

Cyklus je završen vykonáním práce tak, že vítězný agent zašle nové souřadnice, kam se robot přesune. Tím je cyklus ukončen a začíná nové zpracování polohy robota. Pokud se během pohybu robota změni

výrazně hodnoty sensorů je pohyb přerušen a robot zažádá o nové souřadnice.

3 Prostředí

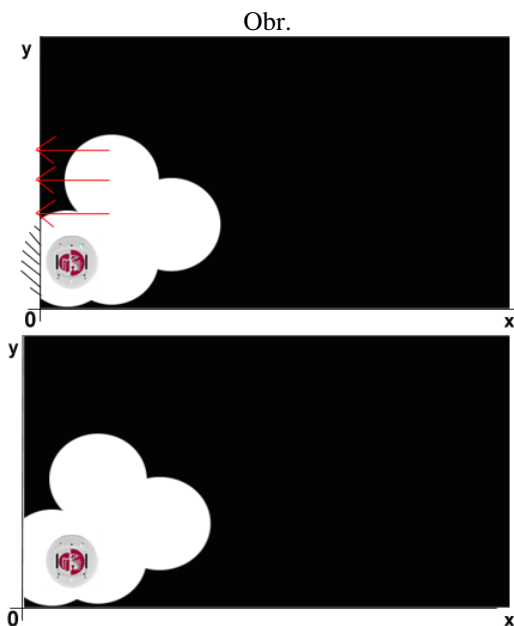
Zpracování prostředí se liší podle toho v jakém režimu naše aplikace běží: *Simulátor/Robot*.

Jestliže nebudeme používat fyzického robota a zvolíme možnost simulátoru, bude naše prostředí reprezentováno velice jednoduše. Jednoduše z toho důvodu, že není potřeba, aby se robot učil celé prostředí. Zkrátka bude mít možnost nahlédnout do vytvořeného pole se souřadnicemi námi umístěných překážek. Robot bude simulovat senzory vytvářením kružnic kolem sebe. Po každém prohledávání svého okolí se budou testovat jednotlivé souřadnice překážek, zda leží uvnitř nebo vně kružnice o předem daném poloměru, jenž by měl simulovat dosah reálných sensorů. K ověření zda bod překážky leží uvnitř nebo vně kružnice využijeme jednoduchou pythagorovu větu. Po ověření se vytvoří výsledný řetězec, který se dále posílá manažerskému agentu.

Pokud zvolíme reálného robota bude reprezentace prostředí složitější. Hlavní rozdíl je v tom, že neznáme dopředu rozložení prostoru. Robot se prostor musí naučit sám. K tomu využijeme senzory robota. Ty nám budou zasílat data o okolí robota a z těchto dat budeme plnit naše pole překážek. Pokud se prostor změní bude se měnit i pole překážek.

Dalším velice důležitým rozdílem je nedeterministická velikost prostředí. Pokud používáme simulátor rozměry prostředí, ve kterém se robot pohybuje, se nemění stejně jako pozice překážek. Avšak u reálného prostředí neznáme předem rozměry a ani uspořádání předmětů v prostoru. To nám ztíží použití soustavy souřadnic, avšak pouze do jisté míry.

Soustavu souřadnic budeme i nadále používat, ale prostředí uděláme dynamičtější. Budeme nastavovat osy x a y podle hranic prozkoumaného prostoru. Osy x a y budou znázorňovat vždy hranici zmapovaného prostoru. Pokud by robot zjistil, že prostor pokračuje za zmapovanou oblastí, osy x a y se přenastaví tak, aby nová oblast tvořila hranici a veškeré dosud získané body se přepočítají, aby odpovídaly novým hodnotám viz. Obr. 3.



Tím vyřešíme libovolnou velikost prohledávaného prostoru a můžeme i nadále používat soustavu x, y souřadnic.

4 Emergence

Jev, který známe jako emergence, nejde až tak jednoduše popsat. Je to jev, který vzniká v nižších vrstvách systému, kde jej není možné dobře identifikovat a jenž postupuje výše, kde se tento jev projevuje vizuálně na makro úrovni systému.

V našem případě může jít právě o pohyb robota, který v nižších vrstvách naší aplikace není řešen, ale samotný pohyb je výsledkem interakce jednotlivých součástí systému.

4.1 Vznik ekonomické krize

V našem systému můžeme pozorovat další projevy zajímavého chování. Jedná se o tzv. *ekonomické krize*. F. A. von Hayek popisuje svůj princip neměnného množství bohatství v ekonomice, jako nástroj omezení rozsahu dopadů těchto krizí. V této aplikaci si toto tvrzení nemůžeme až tak do hloubky ověřit, protože zde nevytváříme složité jednotky, které by ovlivňovali ekonomiku uvnitř systému a které by se zvládali mezi sebou ovlivňovat. Ovšem můžeme se v našem systému setkat se vznikem krizí.

Jak a kdy můžeme tyto krize v systému pozorovat? Představme si, že máme spuštěnou aplikaci v libovolném režimu a s libovolným rozložením aplikací. Robot je ve stavu *prohledávání*. Systém pracuje správně už několikátý cyklus. Již jsou vytvářeny potomci jednotlivých agentů a systém se začíná projevovat pravidelností v přerozdělování práce a odměňování.

Náhle změním stav robota pro příklad ze stavu *prohledává* do stavu *sbírá* nebo *potrava*. Tato změna se začne projevovat v chování. Manažerský agent se snaží dosáhnout cíle stavu robota a na základě toho odlišně vypisuje odměny a jinak vybírá agenty. Tyto změny se začnou projevovat v paralelní aplikaci s agenty. Začnou zanikat generace agentů, které v předchozím stavu byly užiteční a nyní se stávají nadbytečnými. Začínají vznikat nové potomci s vlastnostmi například těch, kteří v předchozím stavu robota nebyli až tak úspěšní. Systém provede reorganizaci tak, aby efektivnost jeho řešení byla opět co nejlepší pro aktuální stav. Po nějaké době se opět systém ustálí a změny nebudou tak viditelné jako při další obměně stavu robota.

5 Závěr

V této práci jsme si ukázali a vysvětlili, jakým způsobem pokračujeme s vytvářením umělé inteligence, která má sloužit k řízení robota Khepera IV.

Pevně věříme, že další výzkum a další pokrok nás přivede k zajímavým závěrům, jenž by mohly pomoci při vytváření umělého života.

Již nyní jsme byli schopni identifikovat v systému jevy, o kterých jsme se v této práci zmínili, jako například ekonomické krize.

Otázkou pro nás nadále zůstává, zda tyto doprovodné jevy, které mohou vznikat v složitých systémech a které vznikají i v reálném lidském světě, nejsou argumentem, k dalšímu zkoumání.

V tomto závěru bych chtěl upozornit na jednu teorii, která nás napadla během pozorování chování našeho systému. Pokud vezmeme libovolný přírodní nebo lidský systém např. ekonomika, ekosystémy (voda, vzduch, fotosyntéza atd), doprava atd., kde se na fungování daného systému podílí menší samostatně fungující jednotky (živé organismy, vhodné agenty), není možné tímto způsobem získat super organismus?

Kde samotný super organismus by byl emergentním jevem vybraného přírodního nebo lidského systému nebo spoluprací více takových systémů. Pokud se na tento problém podíváme z většího nadhledu, mohlo by to znamenat, že i sama planeta je živý organismus a vše co na ni žije vytváří planetární super organismus.

Pokud by tomu tak bylo, pak systém, který nyní vytváříme, by se mohl po dalším zkoumání stát nejen umělou inteligencí, ale také umělým životem. Pro mnoho dalších výzkumníků by tato aplikace, tento výzkum mohl určovat směr k dalším kroků výzkumu umělého života, ale i života jako takového.

Literatura

Mamula, L. (2013). *Simulace hněvu v Hayekově multiagentovém systému*. Diplomová práce, Opava.

Kubík, A. (2004). *Inteligentní agenty*. Computer Press, a.s. Brno. ISBN: 80-251-0323-4

Spell B. (2002). *Java Programujeme profesionálně*. Computer Press, a.s. Praha. ISBN: 80-7226-667-5

Thagard P. (2001). *Úvod do kognitivní vědy. Mysl a myšlení*. Portál, s.r.o. Praha. ISBN: 80-7178-445-1

Petrů M. (2007). *Fyziologie mysli*. Triton. Praha. ISBN 978-80-7254-969-6

Mamula L. (2014). Fuzzy logika v multiagentovém systému s umělou ekonomikou. *Kognitivní věda a umělý život*. 2: 135-139.

Duží M. (2012). *TIL jako procedurální logika*. Aleph. Bratislava. ISBN 978-80-89491-08-7.

Cognitive stress during driving task

Ján Marek, Marek Otáhal, Olga Štěpánková

České vysoké učení technické v Praze
Karlovo náměstí 13, 121 35 Praha 2, Česká republika
{[marekj23](mailto:marekj23@fel.cvut.cz), [otahama2](mailto:otahama2@fel.cvut.cz), [stepanko](mailto:stepanko@fel.cvut.cz)}@fel.cvut.cz

Abstract

Our study is focused on computerized cognitive training (CCT) as a means of improving individual's cognitive abilities, esp. in reducing cognitive decline in elderly patients. According to psychological research, solving meaningful and challenging tasks is more effective than solving puzzles or artificial tasks in cartoon world. That is why we have decided to challenge our users in complex driving situations. We review existing driving simulators and extend one to provide a featureful, yet simple to use and affordable driving simulator & experiment environment. Our solution is available as open source at github.com/breznak/openDS/ and offers convenient and reproducible way for interested researchers who plan to set up their own experiments using designed “cognitive distractors” as sources of stress and recording the driving situation, cognitive load and (optionally collected) biosignals in real-world locations - familiar to the user, locations.

Keywords: driving, stress, cognitive load, cognitive training, elderly, simulator

1 Introduction

Recent studies confirm that computerized cognitive training (CCT) programs can have positive impact on the cognitive performance of adults in some cognitive tests including memory, speed of information processing and visuospatial skills [1]. One of the mundane tasks that naturally challenge human cognitive abilities is car driving: which cannot be performed successfully without ability to coordinate several diverse activities and accomplish multiple goals simultaneously, ie.to ensure multitasking. It is tempting to apply a driving simulator as a basis of CCT focused to multitasking performance. NeuroRacer has been applied for that purpose in a pilot study [2] with 174 participants involved. Authors of this study claim that “. a custom-designed video game can be used to assess cognitive abilities across lifespan, evaluate underlying neural mechanisms, and serve as a powerful tool for cognitive enhancement”. Their results indicate that an adaptive, interference-rich, video game approach can serve as a therapeutic tool for diverse populations suffering from various forms of cognitive deficits.

What should be the main features of a video game to be used in this context? Certainly, it should

- be highly customizable to match abilities and limits of its user ,
- be scalable for many degrees of complexity of the considered tasks as well as to number of tasks required from the user simultaneously,
- be open to further modifications according to the design of researchers who work with.

This paper reviews the design choices we have taken to create such a game that can be performed in a very realistic setting. Special attention is given to different sources of distraction that can appear during driving and that can be perceived as extra load in multitasking test or even as stress factors. In order to identify changes in mood or occurrence of stress in the user as soon as possible the game software supports collection of user's physiologic data from body area network (BAN) sensors that can be selected according to the requirements of the test designer.

Cognitive functions fundamentally affect how a person behaves. Cognitive functions are in fact processes we use to recognize and know world around us, react to events of everyday life, adapt to different situations and learn new things as result of experience or interaction with other people. Decline of cognitive functions can be caused by number of factors including stress [3] or aging [4]. Well-designed cognitive training is currently promoted as an effective instrument for prevention of these processes [5]. We assume that sessions on a realistic driving simulator can act as an attractive instrument for complex training of adult drivers. To verify this hypothesis we need to get access to such a computer simulator. With improvement of technology and graphic in games we are able to deliver realistically looking environment of simulation and create various stress situations which might lead to improvement of cognitive behavior and prepare users for real life situations.

2 Cognitive functions, decline and training

Cognitive functions are various processes including memory, ability to concentrate, visuospatial orientation abilities, ability

to understand information, decide and plan to the future, deal with problems and communicate using language [6].

Cognitive system is performance relies on complex orchestrated integration of several consecutive functions, which are related to each other and cannot work correctly as standalone. An activity to keep these functions working properly, so called cognitive training, is becoming more important at higher age [7]. Derived from research, it is possible to exercise brain similarly as muscles of human body [5, 8] and this is considered as very important for elderly people who have bigger risk of forgetting information [9]. Cognitive training is very important as prevention for people who are often exposed to different risk factors like stress, small amount of sleep, alcohol and cigarette use or lack of social interactions.

Main methods of brain exercise are mental exercise as an act of performing mentally stimulating task beneficial to improve and enhance memory, multitasking – various areas of brain are stimulated by doing many different tasks after each other or at once without brain getting tired, meditation and rest, creative activities as painting writing etc. and finding and learning new information.

3 Stress

Stress is feeling or state of mind experienced by an individual who is pressed by and/or exposed to situations and demands exceeding his/her personal skills and/or social resources he/she can mobilize [10]. In today's world stress plays a prominent role in daily life, evoking negative associations [11]. It has been widely confirmed using different methods over decades that stress is associated with heart disease, immune system problems, headaches and lack of sleep [12, 13, 14].

3.1 Stress and cognitive decline

There are many evidences that cognitive performance changes when one is under stress situation. In the short term, acute stress can be helpful and serve as enhancement of some types of mental functioning. However, chronic and unrelieved stress can produce many negative cognitive changes including memory loss, problems with proper thinking, concentration and behavior. [15] Aside from that stress is documented as an important risk factor which can contribute to developing of Alzheimer's disease and other dementia [16].

3.2 Stress measurement

Once described in the field of psychology, stress became concept that can be measured by many different ways. These include psychological questionnaires covering wide range of psychological symptoms that can be invoked by chronic stress exposure or physiological measures such as blood pressure,

vagal tone or salivary alpha-amylase and cortisol measurement [17].

3.3 Stress in driving and traffic

Experienced drivers perceive driving as a daily routine, involving sequential maneuvers such as maintaining a steady speed, changing lane, breaking and pulling away at the lights. Even these events which could be considered trivial can create some kind of stress level. This level is significantly increased by adding frequent and unpredictable dangerous or even frightening factors to seemingly harmless traffic situations like being followed too closely by another vehicle, driving on unknown path, lack of signaling, having to break hard to avoid collision or getting distracted by environment. Such situations are often present in daily driving routine thus leading into conclusion that driving can be one of the highly stressful activities [18, 19, 20].

This paper first reviews how current driving simulators can mimic considered situations. This is done with intention to suggest features that should be added to create from the resulting driving simulator a useful instrument for studying stress factors related to driving, for measuring cognitive load and improving memory and other cognitive functions by playing car driving simulation games.

4 Driving simulators

Driving simulators are devices or programs providing various driving scenarios which are as close to real life situations as possible. They are used in training of future driver's education courses, private businesses or entertainment. Big role of driving simulators is also medical and human factor monitoring driver performance, attention and behavior. Simulators are also often used in car industry to develop better driver assistance systems or evaluate new vehicles.

4.1 Driving simulators as computer programs

Driving simulators as software can be divided into two main categories, racing simulation and training simulation. Racing simulation or sim racing attempts to accurately simulate car racing, typically with real-world variables such as fuel usage, engine or tire damage and suspension settings. To be competitive in car racing simulator one must understand all aspects of car control. On the other hand training simulators are more beginner friendly as they are focused on learning and enhancing basics of driving.

4.2 Open source driving simulators

Our review is focused mostly on the open source driving simulators which come with affordable price for developers and users. With this option one can provide a simple and cheap

solution available for use in people's homes without purchasing a big simulator. Our work with open source simulator will also be transparent and anyone can benefit from our contributions.

4.2.1 TORCS

The Open Racing Car Simulator ([TORCS](#)) is a 3D car racing simulator designed to enable pre-programmed AI drivers racing against one another and on the same time allowing user to control vehicle using keyboard, mouse or wheel input. The TORCS Racing Board hosts website competition where users upload artificial intelligence controlled “robots” to compete thus doesn't require players to compete in real-time [21].

4.2.2 VDrift

[VDrift](#) is a car racing simulator with focus on drift racing. Its goals are to be high-quality open source driving simulator with challenging game-play and very accurate realistic vehicle physics. Some features are simulation-grade driving physics, real-world based tracks, different kind of gaming input or variable difficulty on AI [22].

4.2.3 OpenDS

[OpenDS](#) is a driving simulator developed primarily for scientific purposes, with core components created at the German Research Center for Artificial Intelligence. Simulator is written using JMonkeyEngine. OpenDS features realistic engine model, simple traffic simulation, traffic light, various weather conditions and support of driving task and environment design [23].

4.3 Simulator evaluation

Each of the upper mentioned simulators targets different type of users. OpenDS simulator proved to be the best choice for our project to simulate distraction factors for current drivers because it has been designed as primary scientific driving simulator providing various options for data analyzing and hardware support.

5 Program implementation

5.1 Goals of our project

In our project we want to

- offer a realistic virtual representation (map) of real world streets and cities,
- design disruptive elements or tasks that the user must handle/pay attention to during driving (so called "distractors"),

- create a scenario editor, that lets the experiment designer or a user choose which distractors and how often (probability) will appear,
- design a function that provides a score of cognitive complexity of a given situation.

5.2 Hypothesis

We believe that

- cognitive functions can be enhanced by playing realistic game with the goal to complete attention/focus demanding tasks,
- driving through already visited and recognized locations can improve success rate in simulation results,
- the user can gain bigger confidence for some real life situations by mastering in advance the details of the journey he/she plans to accomplish but which he/she does not know well using the simulator (learning and practicing on the planned unknown road in the simulator can lower the stress levels created by visiting such place for the first time in reality and thus make it easier for the driver to reach the destination safely).
- idea of participating in CCT with car simulator can be made more attractive for a senior user if he/she is invited to ensure some tasks that are truly familiar to them (journey from home to the office) using the driving simulator because they will feel more involved and accomplishing any of the tasks will support their self-esteem (and memory recall).

5.3 OpenStreetMap

Our project uses [OpenStreetMap \(OSM\)](#) to capture a real-world interpretation of road or city in where user plays the simulation [24]. Program can be also used to train driving on a yet unknown road in order to prepare oneself for a future real life experience on a real road. With using OpenStreetMap each user can easily personalize the game and experience driving through familiar environment. The resulting game will be then perceived by its users primarily as a challenge offering some additional benefits.

At current stage, our project is unable to process too large selections of a map from OSM, because of rapid frame rate drop.

5.4 Options menu

After program starts and Distraction driving task is selected from the Driving task selection menu an options menu appears. In option menu various driving distractors offered in the simulation can be selected. Another option allowing user to set specific probabilities for corresponding distractors to appear in preset amount of time is also present here. Simulation starts running after selecting start button.



Fig. 1: Options menu how it appears after running the program.

5.5 Distractors

Distractors are different distraction variables / objects and effects appearing in every simulation after selected in options menu. Each one has its own probability to appear, which can be also set. Distractors causing stimulation or stress situation for the user can be related to phone calling, outside objects, texting etc. [25]. We would like to use wide variety of distractions to simulate real-world situations as best as possible.

5.5.1 Weather distraction

Weather distraction is environment type of distraction generating rain or snow to drop from the sky or fog to appear in the simulation. This type of distractions creates harder visibility conditions for the driver. When selected weather distraction appears in simulation until the end of session.

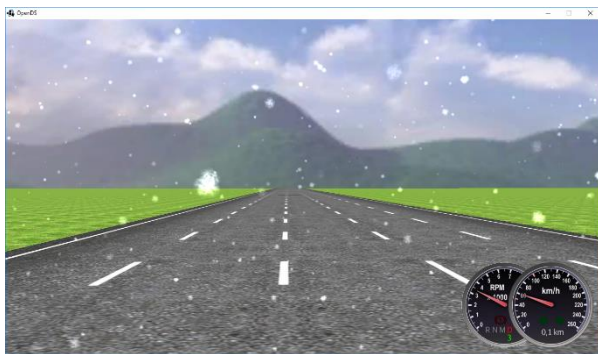


Fig. 2: Weather distraction (snow) showing in the simulation.

5.5.2 Dropping box distraction

Dropping box is a type of distraction resulting from an object (box) that falls from the sky in front of the car into variable distance. Box is then physically present in the world and

represents a danger for the driver who can interact with it by hitting it.

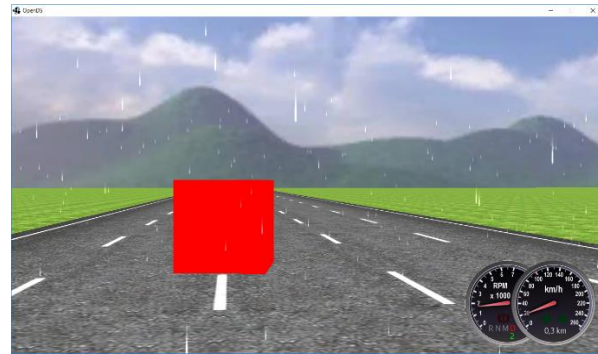


Fig. 3: Box distraction appearing in the simulation while weather distraction is set to rain.

5.5.3 Pedestrian distraction

Pedestrian distraction creates pedestrian-type of the box next to the car which acts like a walking person. Pedestrian can also appear in front of the car wanting to cross the street. Car must then stop to avoid collision.

5.5.4 Sound distraction

Sound distraction when triggered produces prerecorded sound file which can be recording, music (radio like) or plays sound of car horn. This might create stressful situation to control car with even more distractors present in particular situation. Played sound can be changed by changing sound file in Assets folder to meet better the requirements of the user or the experiment.

5.5.5 Darkening distraction

After preset time Darkening distraction generates fog to appear in simulation. This type of distraction is unique type of weather distraction and does not change behavior if fog as weather distraction is not selected. Unlike weather distraction fog, the darkening distraction will remain present only for preset amount of time and disappear afterwards.

5.5.6 Text distraction

Text distraction generates questions that appear in middle of the screen of the simulation. These questions are selected from the text file which can be modified by the tester or user. Text file can contain questions about current events in the simulator such as "How many boxes did you see so far" or "What color was the last box". These questions appear in random order. After user types in his answer and clicks corresponding button, answers are compared with the preset right answer value given

in the text file and saved into a different text file. This feature allows to create environment where user must obtain and retain information under stress conditions - while performing main driving task with distracting elements present. While user is answering question, driving simulation is paused.

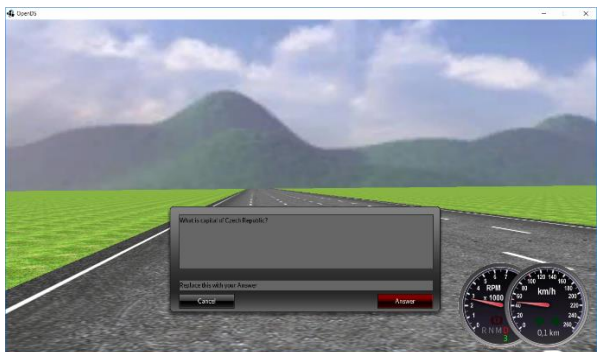


Fig. 4: Text distraction appearing in the simulation while weather distraction is by default clear.

5.6 Cognitive load function

While simulation is running, the cognitive load function provides a score for each point in time. Its purpose is to evaluate the driving situation (environment) in a specific moment and generate a score.

A research of cognitive load and distraction in [25, 26, 27] suggests that common sources of stress are:

- amount of distractors selected,
- amount of distractors appearing in that moment on the screen,
- speed of the users vehicle,
- knowledge of the simulated road (e.g. shape/complexity of a crossroad) to determine how difficult it is to complete the task in an unknown environment,
- all objects surrounding the car in the moment, such as: different cars, objects on or next to the road, pedestrians and others.

Results from the simulation are then stored for later evaluation.

Conclusion

Our research focuses on creating realistic driving simulator as a means to enhance and study cognitive functions of elderly people. We want to use a controlled stressful environment, created by different types of distraction, to achieve real-world conditions drivers meet while performing driving task. A real-world representation of street also creates familiar environment for the users. Many distraction elements are designed to stimulate memory. We managed to design various suitable sources of distraction during the observed driving task and

implement them in a simulator providing a suitable working environment for our planned experiments.

Main features of our experimental driving platform include:

- various stress factors ("distractors") that can be applied in a reproducible manner,
- real-world environments (maps) that can offer training in familiar, or on the contrary new-and-complex locations,
- convenient graphical interface for experiment designers and users,
- logging functionality used to evaluate the whole driving-time experience by the researcher,
- customizable driving simulator available for free and on all PC platforms.

Future research

With the simulator successfully implemented, we wish to proceed to run a battery of experiments and evaluate the observed recordings or driving situations (recorded cognitive score) vs. success rate when dealing with the stress situations, as suggested in upper mentioned hypothesis.

In future research we want to enhance our project by implementing better AI for pedestrians and implement "car opponents". We also want to develop reliable scenario-creating tool for better and easier implementation of real-world environment and further enhance existing and new distraction elements.

Acknowledgements

This work has been supported by the CVUT grant no. **SGS16/231/OHK3/3T/13**.

We would like to thank our colleagues from Agent Technology Center (ATG), Jiří Vokřínek and Martin Shaeffer, who worked on OpenDS project already [28] and provided their valuable expertise that greatly assisted our research.

Literature

- [1] [Laver K1](#), [Hamilton C2](#), [McCluskey A3](#).: Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. [Aust Occup Ther J](#). 2015 Jun;62(3):223-4
- [2] Anguera J.A. et al.: Video game training enhances cognitive control in older adults. *Nature*, vol. 501, pp.97-101, September 2013
- [3] [Marieke S. Tollenaar](#), [Bernet M. Elzinga](#), [Philip Spinhoven](#), [Walter A.M. Everaerd](#). The effects of cortisol increase on long-term memory retrieval during and after acute psychosocial stress. University of Leiden, Section of

- Clinical, Health and Neuropsychology, Wassenaarseweg 15, 2300 RB Leiden, The Netherlands, March 2008.
- [4] Piperhoff P et al. Deformation field morphometry reveals age-related structural differences between the brains of adults up to 51 years. *J Neurosci*. 2008 Jan 23;28(4):828-42.
- [5] [Susanne M. Jaeggi](#), [Martin Buschkuhl](#), [John Jonides](#), and [Priti Shah](#). Short- and long-term benefits of cognitive training. Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1043. 2011
- [6] Sternberg, R. J., & Sternberg, K. (2009). *Cognitive psychology* (6th Ed.). Belmont, CA: Wadsworth, Cengage Learning.
- [7] "Cognitive training" https://en.wikipedia.org/wiki/Cognitive_training Accessed on 23. 4. 2016
- [8] Karlene Ball, PhD et al Effects of Cognitive Training Interventions With Older Adults. *JAMA*. 2002;288(18):2271-2281. doi:10.1001/jama.288.18.2271.
- [9] [Pilar Andrésa](#), [Martial Van der Linden](#) & [Fabrice Parmentier](#). Directed forgetting in working memory: Age-related differences. *Memory*: Vol 12. Issue 2, 2004. Published online: 24 Jun 2010.
- [10] George Fink. *Stress Science: Neuroendocrinology* 2010 ISBN: 9780123750662.
- [11] Boundless. "Defining Stress." Boundless Management. Boundless, 20 Jan. 2016. Retrieved 30 Apr. 2016 available at: <https://www.boundless.com/management/textbooks/boundless-management-textbook/organizational-behavior-5/stress-in-organizations-43/defining-stress-223-7290/>.
- [12] James S. House. Occupational Stress and Coronary Heart Disease: A Review and Theoretical Integration. *Journal of Health and Social Behavior* Vol. 15, No. 1 (Mar., 1974), pp. 12-27.
- [13] DeLongis, Anita; Folkman, Susan; Lazarus, Richard S. The impact of daily stress on health and mood: Psychological and social resources as mediators. *Journal of Personality and Social Psychology*, Vol 54(3), Mar 1988, 486-495.
- [14] Selye, Hans. *Stress in health and disease*. Butterworth-Heinemann, 2013. ISBN 0-407-98510-7
- [15] Cavanagh, J. F.; Frank, M. J.; Allen, J. J. B. (7 May 2010). "Social stress reactivity alters reward and punishment learning". *Social Cognitive and Affective Neuroscience* 6 (3): 311–320.
- [16] Lena Johansson, Xinxin Guo, Tore Hällström, Maria C Norton, Margda Waern, Svante Östling, Calle Bengtsson, Ingmar Skoog. Common psychosocial stressors in middle-aged women related to longstanding distress and increased risk of Alzheimer's disease: a 38-year longitudinal population study. *BMJ Open* 2013;3:9 e003142 doi:10.1136/bmjopen-2013-003142.
- [17] Fernand-Seguín, How to measure stress in humans? Centre for Studies on Human Stress; Hospital Quebec, Canada 2007.
- [18] Joshua D. Clapp, Shira A. Olsen, Sharon Danoff-Burg, J. Houston Hagedorn, Edward J. Hickling, Vivian S. Hwang, J. Gayle Beck [Factors contributing to anxious driving behavior: The role of stress history and accident severity](#). Published in final edited form as: *J Anxiety Disord*. 2011 May; 25(4): 592–598.
- [19] Aaron S. Baker, Scott D. Litwack, Joshua D. Clapp, J. Gayle Beck, Denise M. Sloan. [The Driving Behavior Survey as a Measure of Behavioral Stress Responses to MVA-Related PTSD](#). *Behav Ther*. Author manuscript; available in PMC 2014 July 9. 45(3): 444–453.
- [20] James, Leon and Nahl, Diane (2002). *Dealing With Stress And Pressure In The Vehicle*. Taxonomy of Driving Behavior: Affective, Cognitive, Sensorimotor. University of Alberta Press, Edmonton, Canada.
- [21] "TORCS", <https://en.wikipedia.org/wiki/TORCS>, accessed on 28. 4. 2016
- [22] "VDrift", <http://vdrift.net/>, accessed on 28. 4. 2016
- [23] "OpenDS", <https://www.opens.eu/software/features>, accessed on 28. 4. 2016
- [24] "OpenStreetMap", <https://www.openstreetmap.org/>, accessed on 27.4.2016
- [25] Kristie L. Young, Paul M. Salmon. Examining the relationship between driver distraction and driving errors: A discussion of theory, studies and methods. Monash University Accident Research Centre (MUARC), Building 70, Monash University, Clayton, Victoria 3800, Australia.
- [26] [John Sweller](#). Cognitive load theory, learning difficulty, and instructional design. University of NSW, Australia 2002
- [27] Fred Paas, Juhani E. Tuovinen, Huib Tabbers & Pascal W. M. Van Gerven. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist* Volume 38, Issue 1, 2
- [28] Martin Schaefer, Jiri Vokrinek. Developing Agent-Based Driver Assistance Systems Using AgentDrive. *Advances in Practical Applications of Agents, Multi-Agent Systems, and Sustainability: The PAAMS Collection* Volume 9086 of the series Lecture Notes in Computer Science pp 312-315. 21 May 2015

Aproximální numerický systém u dětí

Michala Plassová^a, Michal Vavrečka^{a,b}, Michael Tesař^a

^aJihočeská univerzita v Českých Budějovicích
Dukelská 9, České Budějovice 1

^bCIIRC, CVUT, Technická 2, Praha 6

mplassova@pf.jcu.cz, vavrecka.michal@gmail.com, michtesar@gmail.com

Abstrakt

Aproximální numerický systém (ANS) lidé (ale i další živočišné druhy) využívají v situacích, kdy je potřeba v okamžiku odhadnout počet prvků množiny, porovnat velikost dvou a více množin nebo odhadnout vzdálenost mezi objekty. V našem výzkumu se zabýváme popisem neurálních korelátů ANS u předškolních dětí. Na základě získaných dat zjišťujeme, že velkou roli v úspěšném řešení předkládaných kognitivních úloh nehrají numerické kompetence v pravém slova smyslu, ale schopnost probanda vizuálně vnímat jednotlivé prvky množiny jako plošný celek. Snažíme se o revizi a reformulaci aproximálního numerického systému, jakožto i redefinici možností, jak může být tento systém využit v edukačním procesu.

1 Úvod

Zájem odborníků o porozumění neurálním korelátům matematických schopností není v literatuře novým tématem, naopak první studie zabývající se spojitostí mezi patologickým poškozením nervové soustavy a dopady na matematické funkce lze datovat do počátku 20. století. V roce 1908 Lewandowsky a Stadelman publikovali vůbec první report o narušení matematických schopností v důsledku poškození mozku. Zhruba o dvě dekády později pak Henschen (1920) poprvé použil termín „akalkulie“, aby pojmenoval pokles kognitivní kapacity pro počítání v důsledku poškození mozku. Na začátku nového století Dehaene (2011) umístil obecné matematické schopnosti v mozku do oblasti, která leží za centrální brázdou - do parietálního (temenního) laloku. Přesněji řečeno se zde nachází centrum pro tu část matematických schopností, kterou označujeme jako aritmetika a s již jsme se setkali na základní škole, když jsme se učili sčítat, odčítat, násobit a dělit.

Bližší charakteristika ANS bývá nejčastěji popisována jako hrubá vnitřní reprezentace čísla (Gallistel & Gelman, 2000; Feigenson et al., 2004), která není závislá na jazyku nebo symbolické matematice (Xu, Spelke, 2000; Gordon, 2004; McCrink, Wynn, 2007; Spaepen et al., 2011). Existují přesvědčivé důkazy (Feigenson et al., 2004; Halberda, Mazocco, & Feigenson, 2008; Gilmore, McCarthy, & Spelke, 2010; Mazocco, Feigenson, & Halberda, 2011b; Piazza et al., 2010; DeWind & Brannon, 2012; Park, Brannon, 2014) o tom, že jde o klíčový systém tvořící základy symbolických matematických dovedností u dospělých lidí.

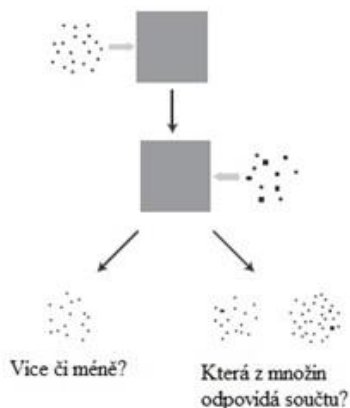
Z neuroanatomického hlediska můžeme ANS lokalizovat do oblasti intraparietální brázdy (intraparietální sulcus – IPS) (Dehaene et al. 2003; Cohen Kadosh et al., 2008), kde je ohraničená inferiorním (IPL) a superiorním (SPL) parietálním lobem (Dehaene et al., 2003; Dehaene, 2011; Nieder, 2013). Prioritní jsou pro nás výsledky mnoha studií (např. Isaacs et al., 2001; Rotzer et al., 2008; Rykhlevskaia et al., 2009; Mussolin et al., 2010), které uvádí, že jak struktura, tak i neurální aktivita v této oblasti odráží interindividuální rozdíly v matematických dovednostech. Zjednodušeně pak množství a kvalita neurálních spojů v této oblasti pozitivně koreluje s úrovní obecných matematických dovedností. Naše dlouhodobá výzkumná otázka spočívá v ověření možnosti využití stimulace oblasti IPS pomocí ANS úloh za účelem zlepšení obecných matematických dovedností. V aktuálním výzkumu se věnujeme popisu funkce ANS u předškolních dětí. Snažíme se tak jednak omezit případný vliv matematické edukace na strategii, kterou jedinec volí při hrubém odhadu počtu, a zároveň reagovat na fakt, že v literatuře obecně chybí informace o tom, zda ANS podléhá ontogenetickým specifikům.

1.2 Popis a funkce ANS

První složku ANS, jež můžeme pojmenovat jako *aproximativní aritmetiku*, můžeme nejspolehlivěji pozorovat ve chvíli, kdy odhadujeme počet prvků

množiny o více, než čtyřech prvcích. Toto lze ilustrovat na situaci, kdy se na obrazovce počítače na krátkou chvíli objeví skupina černých teček (viz obrázek č.1)¹ a rychle zmizí. Pokud by se v dalším okamžiku objevila další skupina teček a my požádali probanda, aby následně odhadl jejich součet, pak bychom využili funkce aproximální aritmetiky.

(A) Aproximální aritmetika



Obrázek 1 Úlohy ANS, aproximální aritmetika

Druhou složkou je senzitivita ANS vůči numerické vzdálenosti rozsahu, někdy též nazývaná jako aproximální porovnávání počtu. Numerickou vzdáleností se v tomto případě rozumí absolutní rozdíl mezi dvěma kontrastními množinami. Například reakční čas odhadu je kratší a přesnost je vyšší, když posuzují, zda je více 12 bílých teček oproti 3 černým (viz obrázek č. 2)², než když posuzují, zda je více 6 bílých teček oproti 3 černým (Haist et al., 2015).

Aproximální porovnávání počtu



Obrázek 2 Úlohy ANS, aproximální porovnávání počtu

Podle některých autorů (Halberda, Mazocco, & Feigenson, 2008; DeWind & Brannon, 2012; Roitman et al., 2012) se přesnost odhadu a reakční čas při zapojení ANS řídí Weberovým zákonem (in Plháčková, 2007), který tvrdí, že intenzita smyslového vjemu je logaritmičtě závislá na intenzitě fyzikálního podnětu. Použijeme-li opět příklad s černými a bílými tečkami, pak by se Weberův zákon (tamtéž) manifestoval tím způsobem, že čím větší je početní rozdíl mezi černými a bílými tečkami, tím rychlejší je reakční čas a vyšší přesnost. Jsou autoři (např. Sasanguie, Defever, Maertens, & Reynvoet, 2013; Sasanguie, Gobel, Moll, Smets, & Reynvoet, 2013; Tibber et al., 2013; Park a Brannonová, 2014), kteří toto tvrzení popírají a argumentují tím, že na každou studii, která logaritmičtý vztah potvrzuje (např. Gilmore, McCarthy, Spelke, 2010; Piazza et al., 2010; Mazocco, Feigenson, & Halberda, 2011a; Mazocco, Feigenson, & Halberda, 2011b; DeWind & Brannon, 2012; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Libertus, Odic, & Halberda, 2012), existuje podobný počet studií (např. Holloway, Ansari, 2009; Inglis, Attridge, Batchelor, Gilmore, 2011; Gobel, Watson, Lervag, & Hulme, 2014), které ho vyvrací. Z tohoto důvodu zaujímáme vůči využití Weberova zákona v oblasti ANS spíše skeptický postoj a domníváme se, že je pro posouzení přesnosti a rychlosti ANS potřeba navrhnout spolehlivý nástroj.

Shrneme-li konečně podstatu zmíněných funkcí ANS, pak můžeme konstatovat, že jsou základem pro vývoj jednoduché aritmetiky. Malý školák se bez toho, aniž by dokázal mentálně zpracovat koncept „více“ a „méně“ a zároveň posoudit numerickou vzdálenost dvou čísel, nemůže naučit sčítat, odčítat, ale ani násobit a dělit.

¹ Převzato z Park, J., Brannonová, E. (2014) Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition* 133 (2014) 188–200.

² Převzato z Park, J., Brannonová, E. (2014) Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition* 133 (2014) 188–200.

2 Metodologie

Nejpálčivějším problémem v současné chvíli zůstává správně zvolená metodologie a detailní studium proměnných, které mohou výzkum v oblasti ANS negativně ovlivňovat. Domníváme se, že klíčová je proměnná v podobě plochy prvků při porovnávání velikosti dvou množin. Negen a Sarnecka (2015) ve svém výzkumu také upozornili na problematiku pochopení termínů „větší“ a „menší“ u předškolních dětí a na vliv velikosti prvků (v jejich případě puntíků) při odhadu počtu. Děti ve věku od 30 do 48 měsíců měly tendenci označovat jako větší tu množinu, která zabrala větší plochu, nikoli tu, která měla větší počet prvků. Z toho vyplývá, že děti se při odhadu velikosti soustředily pouze na velikost plochy, kterou puntíky zabraly. Vychází tak otázka, zda jsou zvolené formulace z lingvistického hlediska správné a pro děti pochopitelné, případně je-li aproximální numerický systém skutečně funkcí početní nebo se jedná o funkci vizuálně-prostorovou. K této otázce nás přivádí fakt, že děti na rozdíl od dospělých nemají naučené žádné sofistikované strategie, jak si s problémem poradit. Zároveň se opíráme o teorii Piageta (1999), který tvrdí, že myšlení předškolních dětí je omezeno na názornou rovinu (tj. ukážeme-li dítěti dvě hromádky bonbónů o 4 a 12 bonbónech, pak bude dítě schopné určit, která hromádka je větší). Nicméně děti nejsou schopny složitějších operací jako jsou kupříkladu konzervace nebo reverzibilita (pokud všech 16 bonbónů vezmeme, nasypeme do široké sklenice a před dítětem přesypeme do sklenice úzké tak, aby „hladina“ bonbónů byla vyšší, pak se dítě bude domnívat, že je v úzké sklenici bonbónů více). Výzkumem funkce a specifík ANS u dětí bychom tudíž mohli získat data o tom, jakou má tento systém podobu v nejnižších ontogenetických stádiích. K tomu je však potřeba důkladně prozkoumat všechny proměnné, které mohou měnit interpretaci výsledků. V našem probíhajícím výzkumu jsme si položili za cíl problematické proměnné identifikovat a prozkoumat.

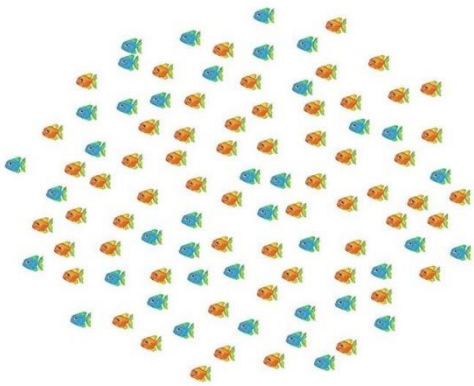
2.1 Participanti

Do výzkumu jsou zapojeny zdravé předškolní děti ve věku od 4.5 roku do 6 let z mateřských škol v okrese Českých Budějovic. V současné chvíli máme data pouze od malého vzorku probandů, a proto v následujícím textu uveřejňujeme výsledky z pilotážních měření, která proběhla na dvou dětských a čtyřech dospělých probandech. Mezi dospělými probandy byli dva rodiče dětí. Dospělé jsme do studie zapojili především s cílem získat zpětnou vazbu na obtížnost

úloh, ale také na vhodnost formulací otázek a zpracování.

2.2 Metoda

Participanty Participanty výzkumu měříme ve dvou oddělených sezeních. Během prvního sezení od rodičů dětí získáváme anamnestická data a u dětí samotných administrujeme Stanford-Binetův inteligenční test. V následujícím sezení děti měříme při řešení námi navržených kognitivních úloh na 64 kanálovém elektroencefalografu (EEG) s aktivními elektrodami. Výzkumný úkol jsme naprogramovali v programu Presentation. Samotná podoba úlohy vychází z výzkumu Parka a Brannonové (2014). Podobu jejich úloh jsme prezentovali v předchozím textu (viz obrázky č. 1 a č. 2). V jistě obměně ale můžeme tento druh úloh najít v celé řadě dalších studií (např. Mazzocco, Feigenson, & Halberda, 2011a; Mazzocco, Feigenson, & Halberda, 2011b; DeWind & Brannon, 2012). V našem výzkumu z metodologických důvodů používáme pouze typ úlohy, ve které je probandovi prezentovaná jedna množina, v níž jsou bílé a černé tečky (viz obrázek č. 2, dolní část). Činíme tak ze dvou základních důvodů. Prvně jsme již zmínili, že předškoláci nejsou schopni složitějších mentálních manipulací (např. konzervace počtu), a proto se domníváme, že typ úlohy v obrázku č. 1 není vzhledem ke své náročnosti vhodný. Dalším důvodem je negativní vliv prostorové architektury úlohy na EEG signál. Kupříkladu pokud na monitoru počítače naráz prezentujeme dvě různé množiny (vpravo bílé tečky a vlevo černé) a po probandovi chceme, aby řekl, kterých teček je víc, pak musí hýbat očima zleva doprava (popř. zprava doleva), aby si obě množiny prohlédl. Oční pohyby mají bohužel negativní vliv prostorové architektury úlohy na EEG signál, a to ve formě šumu (artefaktu), jež signál významně znehodnocuje. Používáme proto rozdělení plochy pomocí středového kříže a prezentované množiny prvků orientujeme na střed. Konkrétní podobu našich vizuálních stimulů ilustrujeme na obrázku č. 3. Jedná se o dětský motiv, který jsme vytvořili specificky dle našich potřeb.



Obrázek 3 Log 1.5

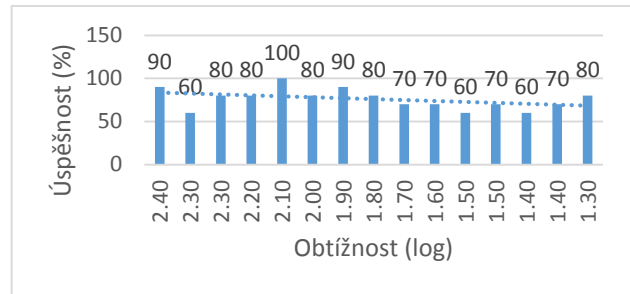
Výzkumné měření je dětem představeno jako hra s podmořskou tematikou, snažíme se tak v dětech navodit zájem a motivaci k řešení úloh. Vizualní stimuly dětem prezentujeme v rychlém sledu (1500 ms na prezentaci stimulu) a přesně kontrolujeme, jakou plochu rybičky zaujmají. Zvolili jsme z toho důvodu konstantní celkový počet rybiček (100 rybiček) s konstantní velikostí. K rozmístění rybiček na ploše používáme Fibonacciho grid. Rybičky se liší pouze zvolenou barvou (oranžové a modré). Pro náš výzkum jsme podle výsledků Parka a Brannonové (2014) stanovili pět základních úrovní poměru zastoupení obou barev rybek (od log 1.5 po log 1.1), nicméně v současné chvíli testujeme i méně náročné varianty (od log 1.9), a to z důvodu, abychom zajistili vysokou úspěšnost a výsledky měření nezkreslovaly náhodné tipy. Neuniklo totiž naší pozornosti, že se u obtížnějších logů (1.3 a méně) úspěšnost snižuje a u některých dětí nepřesáhne hranici náhody (50%). Ačkoli jsme na začátku ANS definovali jako hrubý odhad, pak v těchto situacích nelze s určitostí tvrdit, že děti ANS používají. Z výpovědí dětí vyplývá, že během obtížných logů kupříkladu přemýšlely, kterou barvu rybek již dlouho nestiskly.

Task s dětmi začínáme na základní obtížnosti (1.9) a po vyřešení každých deseti úloh počítač přepočítává, jak si dítě vede, a podle toho mění obtížnost úlohy, tj. při úspěšnosti dítěte > 60% se obtížnost zvyšuje, při úspěšnosti < 50% se snižuje a v rozmezí 51 – 59% se udržuje aktuální obtížnost. V programu Matlab jsme vytvořili generátor, který vytváří randomizované množiny rybek. Ačkoli je celkový počet rybek konstantní, generátor kontroluje, aby se počet modrých a oranžových rybek průběžně měnil dle zvolené obtížnosti a aby se lišilo umístění rybek v prostoru.

3 Dosavadní výsledky

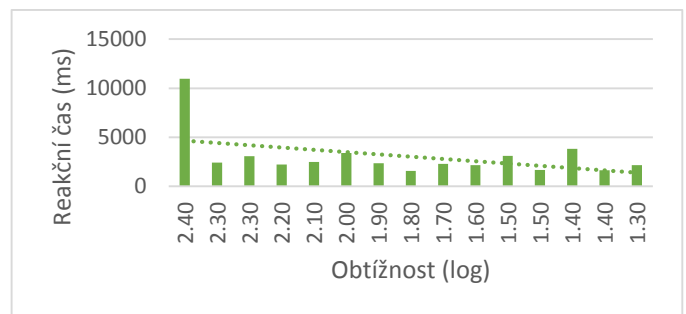
V pilotážní studii jsme si stanovili několik cílů, prvním významným cílem bylo ověřit, zda jsme při snímání předškoláků schopni zajistit kvalitní EEG signál s minimem artefaktů. Následně nás zajímala procentuální úspěšnost a reakční čas (RT) v závislosti na aktuální obtížnosti úlohy.

Získaná data ukazují, že při pilotáži dětští probandi dosahovali vysoké úspěšnosti i při řešení obtížných úrovní úlohy (viz obrázek č. 4)



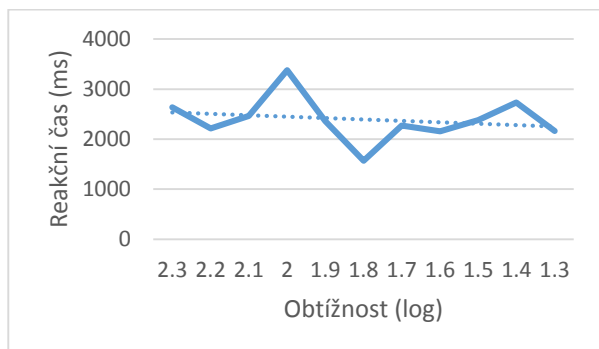
Obrázek 4 Vlivu obtížnosti úlohy na úspěšnost řešení

Zároveň se v křivce zaznamenávaného reakčního času (RT) projeví dva jevy, a to zácvik na začátku měření a následně vliv obtížnosti úlohy na RT (viz obrázek č. 5 a obrázek č. 6). Přičemž na obrázku č. 5 vidíme, že probandé měli nejdelší reakční čas u zácviku (log 2.4), který byl zároveň nejlehčí obtížností. Jev vysvětlujeme jednak samotným zácvikem v úloze a zároveň nutností si zvyknout na způsob zadávání odpovědi pomocí počítačové myši. Následně se RT prodlužoval s narůstající obtížností. U hodnot logu, které se v obrázku č. 5 vyskytují vícekrát (např. log 1.50 a 1.40) probandi kvůli nízké úspěšnosti obtížnost opakovali. Po opakování obtížnosti se RT u druhého pokusu výrazně snižoval. Lze se tak domnívat, že řešením úlohy samotné dochází k učení. Zatím však nevím, zda dochází ke tréninku ANS obecně nebo pouze ke tréninku v našich specifických úlohách.



Obrázek 5 Vliv obtížnosti úlohy na reakční čas.

Na obrázku č. 6 vidíme mediány RT v závislosti na času. Opět je patrná mírná tendence klesání RT v průběhu řešení úloh a narůstání RT u obtížných logů (např. log 1.4). Nicméně je potřeba znovu zmínit, že se jedná pouze o výsledky pilotáže a pro zobecnění (těchto jevů) je potřeba tyto jevy potvrdit u širšího výzkumného vzorku.



Obrázek 6 Průměrné reakční časy napříč úlohami

3.1 Plánovaná analýza EEG signálu a očekávané výsledky

Po získání dostatečného počtu probandů plánujeme porovnávat evokované potenciály dětí s vysokou úspěšností řešení kognitivních úloh s evokovanými potenciály dětí s nízkou úspěšností. Předpokládáme, že nalezneme signifikantní rozdíly v průběhu řešení úlohy, a to zejména v prvních 250 ms, kdy je úloha zrakově zpracovávána. Takové zjištění by podpořilo naši hypotézu, že při řešení předkládaných úloh závisí úspěšnost na vizuálních strategiích.

Dále si klademe za cíl porovnat EEG data s výsledky Stanford-Binetova inteligenčního testu, zejména pak s částí, která vyšetřuje matematické nadání. Očekáváme, že děti, které budou administrované jako nadprůměrné v testu předškolních matematických znalostí, budou úspěšné i při řešení naší úlohy.

4 Budoucí směřování výzkumu

V následujících měření plánujeme měnit velikost rybek tak, aby při odlišném počtu zaujímaly stejnou plochu. Dále nás zajímá případný vliv barvy stimulu (zařadíme stimuly, kde budou počty oranžových a modrých rybek stejné). V neposlední řadě chceme aktivně měnit vzdálenost mezi rybkami a sledovat, zda změny v prostorovém uspořádání budou mít vliv na úspěšnost odhadu a délku reakčního času. Do budoucna zamýšlíme

i průzkum vlivu tréninku ANS funkcí pomocí námi navržených úloh na obecné aritmetické dovednosti.

Poděkování

Tento příspěvek vznikl za podpory Grantové agentury Jihočeské univerzity v rámci grantového projektu 112/2015/S.

Literatura

Cohen Kadosh R, Cohen Kadosh K, Henik A (2008): *When brightness counts: the neuronal correlate of numerical luminance interference*. Cereb Cortex 18:337–343.

Dehaene S (1992): *Varieties of numerical abilities*. Cognition 44:1–42.

Dehaene S (2011): *The Number Sense: How the Mind Creates Mathematics*, Revised and Updated Edition. New York, NY: Oxford University Press.

Dehaene S, Changeux JP (1993): *Development of elementary numerical abilities: a neuronal model*. J Cogn Neurosci 5:390–407.

Dehaene S, Piazza M, Pinel P, Cohen L (2003): *Three parietal circuits for number processing*. Cogn Neuropsychol 20:487–506.

DeWind, N. K., & Brannon, E. M. (2012). *Malleability of the approximate number system: Effects of feedback and training*. Frontiers in Human Neuroscience, 6.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. Trends in Cognitive Sciences, 8(7), 307–314.

Gallistel, C. R., & Gelman, R. (2000). *Non-verbal numerical cognition: From reals to integers*. Trends in Cognitive Sciences, 4(2), 59–65.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). *Symbolic arithmetic knowledge without instruction*. Nature, 447, 589–591.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). *Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling*. Cognition, 115, 394–406.

Gobel, S. M., Watson, S. E., Lervag, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. Psychological Science, 25(3), 789–798.

- Gordon P (2004): *Numerical cognition without words: Evidence from Amazonia*. *Science* 306:496–499.
- Haist F, Wazny JH, Toomarian E, Adamo M. Development of brain systems for nonsymbolic numerosity and the relationship to formal math academic achievement. *Human Brain Mapping*. 36: 804–26.
- Halberda J, Feigenson L (2008): *Developmental change in the acuity of the “Number Sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults*. *Dev Psychol* 44:1457–1465.
- Halberda J, Mazocco MM, Feigenson L (2008): *Individual differences in non-verbal number acuity correlate with maths achievement*. *Nature* 455:665–668.
- Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L (2012): *Number sense across the lifespan as revealed by a massive Internetbased sample*. *Proc Natl Acad Sci USA* 109:11116–11120.
- Henschen, S. E. (1920). *Klinische und anatomische Beiträge zur Pathologie des Gehirns*. Stockholm: Nordiska Bokhandeln.
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children’s mathematics achievement. *Journal of Experimental Child Psychology*, 103(1), 17–29.
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin and Review*, 18(6), 1222–1229.
- Isaacs E. B., Edmonds C. J., Lucas A., Gadian D., G. (2001). *Calculation difficulties in children of very low birthweight: A neural correlate*. *Brain*; 124(Pt 9):1701–7.
- Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, 141(3), 373–379.
- Lewandowsky, M. & Stadelmann, E. (1908). *Über einen bemerkenswerten Fall von Hirnblutung und über Rechenstörungen bei Herderkrankung des Gehirns*. *Journal für Psychologie und Neurologie*, 11, 249–265.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). *Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia)*. *Child Development*, 82(4), 1224–1237.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). *Preschoolers’ precision of the approximate number system predicts later school mathematics performance*. *PLoS ONE*, 6(9), 1–8.
- McCrink K., Wynn K. (2007): *Ratio abstraction by 6-month-old infants*. *Psychol Sci* 18:740–745.
- Mussolin C., Mejias S., Noël M.P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition*;115(1):10–25.
- Negen J., Sarnecka BW. (2015) Is there really a link between exact-number knowledge and approximate number system acuity in young children? *The British Journal of Developmental Psychology*. 33: 92-105.
- Nieder A (2013): *Coding of abstract quantity by ‘number neurons’ of the primate brain*. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 199:1–16.
- Park J., Brannon E. M. (2013). *Training the approximate number system improves math proficiency*. *Psychol. Sci.* 24 2013–2019.
- Park J., Brannon E. M. (2014). *Improving arithmetic performance with number sense training: an investigation of underlying mechanism*. *Cognition* 133 188–200.
- Piaget, J. (1999). *Psychologie intelligence*. Praha: Portál.
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, 116(1), 33–41.
- Plháková A. (2007). *Učebnice obecné psychologie*. Praha, Akademia, vyd. 1.
- Roitman JD, Brannon EM, Platt ML (2012): *Representation of numerosity in posterior parietal cortex*. *Front Integr Neurosci* 6:25.
- Rotzer S., Kucian K., Martin E., von Aster M., Klaver P., Loenneker T. (2008). *Optimized voxel-based morphometry in children with developmental dyscalculia*. *Neuroimage* 39 417–422.
- Rykhlevskaia, E., Uddin, L.Q., Kondos, L., & Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience*, 3(51),1-13.

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2013). The approximate number system is not predictive for symbolic number processing in kindergartners. *The Quarterly Journal of Experimental Psychology*, 1–26.

Sasanguie, D., Gobel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, 114(3), 418–431.

Spaepen E, Coppola M, Spelke ES, Carey SE, Goldin-Meadow S (2011): *Number without a language model*. *Proc Natl Acad Sci USA* 108:3163–3168.

Tibber, M. S., Manasseh, G. S., Clarke, R. C., Gagin, G., Swanbeck, S. N., Butterworth, B., et al. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical predictor of mathematical ability. *Vision Research*, 89, 1–9.

Xu, F., & Spelke, E. S. (2000). *Large number discrimination in 6-month-old infants*. *Cognition*, 74, B1–B11.

Špecifiká kognitívneho výkonu pacientov po ložiskovej ischémii mozgu

Ján Rybár¹, Barbora Cimrová^{1,2}, Igor Farkaš¹, Monika Varga Doležalová³, Roman Rosipal⁴

¹ Centrum pre kognitívnu vedu, FMFI UK, Mlynská dolina, 84248 Bratislava

² Laboratórium kognitívnej neurovedy, ÚNPF SAV, Sienkiewiczova 1, 81371 Bratislava

³ Univerzitná nemocnica Bratislava, Mickiewiczova 13, 81369 Bratislava

⁴ Ústav merania SAV, Dúbravská cesta 9, 84109 Bratislava

Abstrakt

Výskum je súčasťou rozsiahlejšieho projektu zameraného na hľadanie súvislostí spánkových charakteristík a kognitívneho výkonu pacientov po cievnej mozgovej príhode. Na testovanie kognitívneho výkonu sme použili nami vytvorené počítačové verzie troch typov psychologických testov. Test na pozornosť tvorí lateralizovaná verzia (LANT) známeho testu pozornostných sietí s parametrami upravenými pre našu výskumnú vzorku – pacientov prevažne vo vyššom veku. Pracovnú pamäť preverujeme štandardným testom na zopakovanie poradia prezentovaných číslíc v rovnakom a opačnom poradí. Jemnú motoriku stanovujeme pri obkresľovaní predlohy pomocou grafického tabletu. Na vytvorenie referenčných hodnôt sme všetky testy vyskúšali aj na zdravej kontrolnej skupine rovnakého vekového zloženia. Pacienti a zdraví dobrovoľníci sa nelíšili v skóre z testu jemnej motoriky. Avšak, našli sme štatisticky významné rozdiely medzi oboma skupinami v teste pracovnej pamäte, v reakčných časoch v teste LANT vo všetkých podmienkach a tiež v jednej z pozornostných sietí (orienting inhibitory).

1 Úvod

V príspevku sú obsiahnuté parciálne výsledky z rozsiahleho projektu „Zmeny architektúry spánku u pacientov s ložiskovou ischémiou mozgu a ich vplyv na kognitívne funkcie.“

Odhady výskytu kognitívnych poškodení pacientov po cievnej mozgovej príhode sú rôzne vzhľadom na celkovú variabilitu ťažkostí podľa lokality a rozsahu príhody. Mellon a spol. (2015) udávajú podľa súhrnu štúdií 30–50% prevalenciu kognitívneho deficitu u pacientov po príhode, Chen a spol. (2015) hovoria až o vyše 60% prípadov, Moran a spol. (2014) dokonca opisujú príznaky u 15% pacientov po tranzitornom ischemickom ataku (TIA). Väčšina výskumov o kognitívnych zmenách po náhlejšej cievnej mozgovej príhode (NCMP) sa týka ťažších foriem (podľa NIHSS – National Institute of Health Stroke Scale), zatiaľ čo u ľahších foriem sa očakávalo plné uzdravenie a návrat do života (Wolf a Rognstad, 2013). Novšie výskumy

však potvrdzujú pretrvávajúce deficity aj v tejto populácii: exekutívne funkcie u 30–60% pacientov (Wolf, Barbee a White, 2011), celkovo zmenený resp. nešpecifikovane kognitívny status u 39% pacientov (Moran a spol., 2014).

Kľúčovou otázkou v tejto časti nášho výskumu je, či majú pacienti s NCMP rozdielny kognitívny status v porovnaní s kontrolnou skupinou v rôznych experimentálnych podmienkach, a pokiaľ ide o pozornostné systémy, aj v rôznych vizuálnych poliach. Zodpovedať túto otázku sme sa pokúsili prostredníctvom troch psychologických testov.

2 Metódy

2.1 Participanti

Pacienti s NCMP boli do výskumu zaradení na základe nasledujúcich kritérií: majú intaktné fatické (rečové) funkcie, motorický postih v miere, čo im umožňuje účasť na testovaní, neprítomnosť závažnejšej poruchy zraku a sluchu. Do vzorky boli zaradení aj pacienti s úplným ústupom neurotopickéj symptomatiky do 7 dní od vzniku NCMP.

Vzorka vyšetrených pacientov pokrýva všetky základné kategórie NCMP z pohľadu miesta poškodenia: supratentoriálne, cerebelárne a kmeňové. Z hľadiska cieľov nášho výskumu sa zameriavame predovšetkým na pacientov so supratentoriálnym poškodením (t.j. lokalizovaným kortikálne a subkortikálne). Vzhľadom na to, že predmetom skúmania je kognitívny status každej hemisféry, je dôležité, či je poškodená ľavá alebo pravá hemisféra.

Testovanie prebiehalo od 6.10.2014 do 23.2.2016 na vzorke 20 pacientov (16 mužov, 4 ženy), s priemerným vekom 58 rokov (SD = 13). Ručnosť pacientov nebola zisťovaná testom, iba na základe sebahodnotenia. Batéria testov trvala približne jednu hodinu. V niektorých prípadoch prebehol aj retest po troch mesiacoch.

Pokiaľ ide o vzorku zdravých participantov, testy boli vykonávané od 14.7.2015 do 1.4.2016 na vzorke 22 participantov (14 mužov, 8 žien), s priemerným vekom 55 rokov (SD = 10). Retesty vo väčšom časovom odstupe sa neuskutočnili.

2.2 Použité psychologické testy

Keďže základným cieľom projektu je preskúmať súvis charakteristík spánku s kognitívnym výkonom na druhý deň po zobudení, vybrali sme tri základné psychologické testy, u ktorých je predpoklad ich súvisu so spánkom, a to konkrétne na pracovnú pamäť, jemnú motoriku a pozornosť. Test pracovnej pamäte patrí do štandardného portfólia kognitívnych testov. Test jemnej motoriky sme zaradili preto, lebo sme predpokladali veľké rozdiely vo výkonoch medzi pacientmi a zdravými participantmi (čo sa však nepotvrdilo, ako uvidíme ďalej). Najsofistikovanejší je test na meranie pozornosťných systémov (resp. sietí). Za týmto účelom sa zvyčajne používa paradigmatická Posnerova technika s využitím nápovedy (cue) v rôznych podmienkach (Gazzaniga a spol., 2014). Na základe tejto techniky vznikol tzv. ANT (attentional network test), ktorým sa zisťuje nielen behaviorálna evidencia, ale aj neurálne koreláty troch autonómnych pozornosťných sietí – konfliktu medzi cieľovým a distrakčným podnetom, priestorovej orientácie a neustálej pozornosti (alerting) (Posner a Raichle, 1994; Fan a spol. 2002; Posner a Fan, 2013). Neskôr bola vytvorená lateralizovaná verzia testu (LANT), ktorá skúma pozornosťnú kapacitu každej hemisféry, teda hemisferický status každej siete (Greene, 2008).

Test pracovnej pamäte

Ako sme už vyššie naznačili, použili sme štandardný test stanovujúci kapacitu krátkodobej a pracovnej pamäte (digit span), ktorý je súčasťou Wechslerovho testu inteligencie (Kaufman a Lichtenberger, 2006). Počas testu je participantovi postupne prezentovaná séria čísiel (v strede obrazovky), ktorú má za úlohu zopakovať v prezentovanom poradí alebo v obrátenom poradí (kedy je nutné aktívnejšie zapojiť operačnú pamäť). Sekvenciu subjekt zadáva na externej číselnej klávesnici. Po úspešnom zopakovaní správneho poradia sa dĺžka sekvencie zvýši (v opačnom prípade zníži) o jednu číslicu. Kapacita sledovanej zložky pamäte potom zodpovedá najväčšej dĺžke sekvencie, ktorú ešte participant uviedol správne. Tento počítačový test bol implementovaný v aplikácii E-Prime.

Test jemnej motoriky

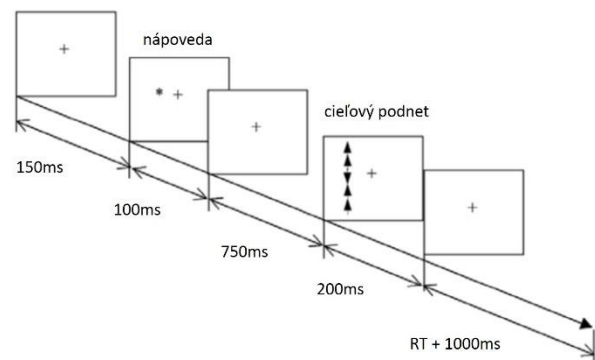
Ako je známe, pacienti s NCMP majú najčastejšie postihnutý motorický systém (Ambler a Bauer, 2010), preto bol zaradený aj jednoduchý test na meranie jemnej motoriky. Pre tento test bol zakúpený tablet s elektronickým perom (zn. Wacom) umožňujúci priamo do počítača zaznamenávať pohyb pera po podložke ako pri klasickom písaní (kreslení). Použitý test jemnej motoriky je založený na prekresľovaní jednoduchých geometrických útvarov (štvorec, trojuholník, zakrivené čiary a pod.) priamo cez predlohu. Skóre sa počíta ako pomer správne prekreslených (prekrytých) pixlov a

celkového počtu pixlov predlohy (pričom pero vytvára čiaru adekvátnej hrúbky).

Pre potrebu snímania a vyhodnocovania vstupu z tohto zariadenia bolo potrebné zvoliť implementačné prostredie odlišné od prostredia použitého pri tvorbe ostatných dvoch experimentov (E-prime). Preto sme sa rozhodli pre vývoj malej aplikácie v jazyku MS Visual C++ s podporou grafickej knižnice SDL a SDL_image. Na obr. 5 sú znázornené príklady použitých predlôh (čierna čiara) už po prekreslení participantom (červená čiara). Predlohu, ako aj pohyb pera a vytváranú čiaru videl participant v reálnom čase na obrazovke počítača.

Test pozornosti

Navrhli sme modifikovanú verziu testu LANT (Greene a spol., 2008), pretože pre náš výskum bolo potrebné upraviť faktory ovplyvňujúce kognitívnu náročnosť (čas zobrazenia, veľkosť a rozstup jednotlivých podnetov, ako aj povolený čas na zaznačenie odpovede) tak, aby bol test vhodný pre našu výskumnú vzorku. Úlohou participantov (či už pacientov alebo zdravých) je detekovať smer cieľovej šípky. Cieľová šípka sa vždy zobrazuje uprostred štyroch distrakčných šípok (teda dvoch nad a dvoch pod cieľovou šípkou (obr. 1). Smer cieľovej šípky je vždy buď v súlade so štyrmi distrakčnými šípkami (kongruentná podmienka) alebo naopak, v nesúlade, keď distrakčné šípky sú orientované v opačnom smere ako cieľová šípka (nekongruentná podmienka).



Obr. 1: Časový priebeh jednej sekvencie v teste LANT. RT = reakčný čas (podľa Green a spol. 2008).

V teste sa používa nápoveda (cue) v tvare hviezdičky. Ak sa objaví na tej istej strane ako cieľový podnet, ide o validnú (správnu) nápovedu, ak na opačnej strane, ide o nevalidnú (nesprávnu, zavádzajúcu) nápovedu. Nápoveda sa tiež môže objaviť na oboch stranách zároveň alebo sa nemusí vôbec objaviť.

Ako vidieť v schéme procedúry testu LANT na obr. 1, po zobrazení fixačného križiku sa môže alebo nemusí zobraziť jeden typ nápovedy, pričom križik zostáva zobrazený neustále. Samotný cieľový podnet sa zobrazí buď napravo alebo naľavo od fixačného

krížika. Participant má za úlohu identifikovať smer stredovej šípky pomocou stlačenia dolného alebo horného tlačidla myši, pričom zrak má upierať neustále na fixačný krížik. Po zaznačení odpovede nasleduje pauza a celá sekvencia sa opakuje.

Oproti pôvodnej verzii testu (Greene, 2008) sme predĺžili čas zobrazenia podnetu zo 170ms na 200ms, samotné podnety sme mierne zväčšili a obrátili sme farebné zobrazenie – v našej verzii sú prezentované biele podnety na čiernom pozadí. Tieto zmeny sme začleňovali postupne počas pilotnej fázy štúdie a kontinuálne sme ich testovali na participantoch nezaraďených v tu prezentovanej vzorke. Hlavným cieľom bolo znížiť nepríjemné pocity (slzenie, bolesť očí, či hlavy) vyvolané silným jasom obrazovky a extrémne drobných podnetov, počas pomerne dlhšej prezentácie pôvodnej verzie celého testu. Naša verzia pozostávala zo štyroch blokov, z ktorých každý trval približne päť minút (aj v závislosti od rýchlosti odpovedí). Po každom bloku bola zaradená prestávka, dĺžku ktorej si participant určil sám.

V našom výskume pracujeme s tromi (resp. piatimi) pozornosťnými sieťami. Všetky sa počítajú ako rozdiel v reakčných časoch (RT) dosiahnutých v dvoch rôznych podmienkach.

1. Sieť neustálej pozornosti (alerting, A) vyjadruje schopnosť využiť upozornenie – krátky rýchly signál (alert) tesne pred zobrazením cieľového podnetu. Sieť A sa počíta ako rozdiel medzi reakčnými časmi s nevýznamovou náповедou (na oboch stranách zorného poľa) a reakčnými časmi bez náповedy.

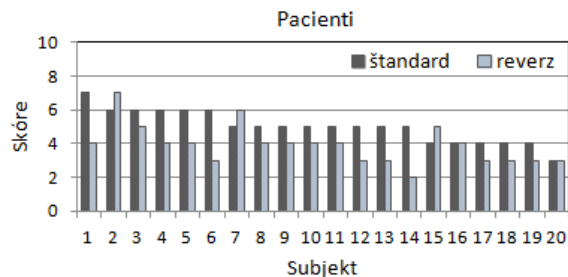
2. Ďalšia pozornosťná sieť je zameraná na konflikt (conflict, C) medzi cieľovým podnetom a distrakčnými podnetmi. Sieť C sa počíta ako rozdiel medzi reakčnými časmi pre kongruentné a pre nekongruentné podmienky (t.j. súhlasné a nesúhlasné smerovanie cieľových a distrakčných podnetov).

3. Vizualno-priestorová pozornosťná sieť (orienting, O) má na starosti zameranie pozornosti na určitú časť priestoru. Sieť O sa počíta ako rozdiel medzi reakčnými časmi v podmienke s valídnu náповедou (zobrazenou na rovnakej strane zorného poľa ako cieľový podnet) a podmienke s náповедou bez významu (zobrazenej na oboch stranách zorného poľa). Táto sieť sa delí ešte na dve podsiete: inhibičnú (O-inhibitory), definovanú ako rozdiel medzi reakčnými časmi v podmienke so zavádzajúcou (nevalídnu) náповедou a podmienke s upozornením bez orientácie (náповeda na oboch stranách zorného poľa) a facilitičnú (O-facilitory), definovanú ako rozdiel medzi reakčnými časmi v podmienke s nevýznamovým upozornením (na oboch stranách) a podmienke so správnou náповедou (na rovnakej strane ako výsledný podnet).

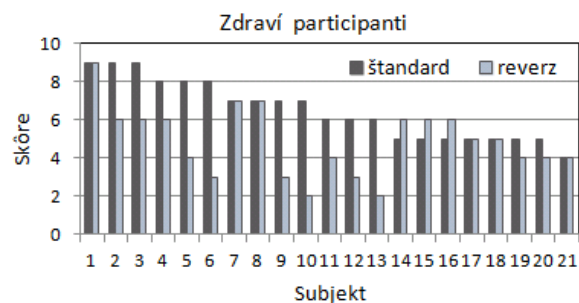
3 Výsledky

3.1 Test pracovnej pamäte

Tento test absolvovali všetci pacienti ($n = 20$). V kontrolnej skupine jeden participant dosiahol minimálne skóre, pravdepodobne nepochopil úlohu, preto bol z analýzy vylúčený ($n = 21$). Skóre subjektov oboch skupín je znázornené na obr. 2 (pacienti) a obr. 3 (zdraví participantí). V oboch prípadoch sú participantí zoradení zostupne podľa ich skóre v štandardne prezentovanom poradí číslíc. Už na prvý pohľad je vidno, že skóre dosiahnuté zdravými participantmi je vyššie.



Obr. 2: Výsledky testu pracovnej pamäte pacientov.



Obr. 3: Výsledky testu pracovnej pamäte v kontrolnej skupine.

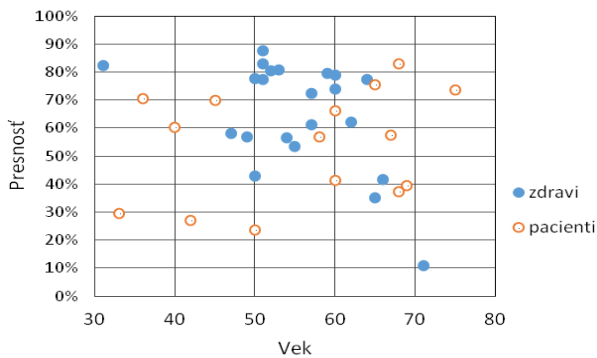
| Typ testu | Zdraví | Pacienti | Signif. |
|------------|-----------|-----------|-------------|
| Štandardný | 6.6 (1.4) | 5.1 (1,0) | $p < 0.001$ |
| Reverzný | 4.9 (1.6) | 3.9 (1.2) | $p = 0.024$ |

Tab.1: Porovnanie výsledkov pacientov s výsledkami zdravých participantov v teste pracovnej pamäte.

Neočakávané bolo, že tak ako v skupine pacientov ako aj v skupine zdravých participantov sa vyskytlo niekoľko výkonov s lepším skóre v reverznom (obrátenom poradí) než v štandardnom variante testu (opakovanie položiek v poradí, v akom boli prezentované). Porovnanie oboch skupín vidieť v tab. 1, podľa ktorej v oboch smeroch zdraví participantí vykazovali vyššie skóre (použili sme smerový t -test pre 2 skupiny s nerovnakou varianciou), pričom väčší rozdiel bol v štandardnom (jeden a pol číslice) než reverznom smere (jedna číslice).

3.2 Test jemnej motoriky

Test jemnej motoriky absolvovala len časť pacientov ($n = 15$) a všetci zdraví účastníci ($n = 22$). Skóre účastníkov v oboch skupinách je na obr. 4. Tab. 2 znázorňuje rozdiel medzi výsledkami pacientov a zdravých účastníkov. Rozdiel medzi skupinami nie je štatisticky významný (nepárový t -test), napriek pomerne veľkému rozdielu stredných hodnôt (no zrejme kvôli vysokej variabilite skóre v oboch skupinách). Možno teda konštatovať, že motorické zručnosti neboli v dôsledku NCMP u pacientov zhoršené v porovnaní so zdravými účastníkmi. Obr. 5 ilustruje niekoľko príkladov motorických odpovedí pacientov.



Obr. 4: Priemerné skóre pacientov a zdravých účastníkov v teste jemnej motoriky.

| Pacienti | Zdraví | Signifikantnosť |
|----------|----------|------------------|
| 54 (19)% | 65 (19)% | $p = 0.1$ (n.s.) |

Tab. 2: Priemerná úspešnosť v teste jemnej motoriky u pacientov a zdravých účastníkov a výsledok t -testu.



Obr. 5: Príklady motorických odpovedí s veľmi malou (11 a 27%) a veľmi vysokou (90 a 84%) presnosťou u pacientov. Červená čiara ukazuje trajektóriu pohybu pera vykonanú účastníkom pri prekresľovaní predlohy (čierna farba).

3.3 Test pozornosti

V prípade oboch skupín sme počas testu LANT zaznamenávali reakcie účastníkov na podnety separátne v šiestich podmienkach: kongruentné alebo nekongruentné smerovanie distrakčných šípok, validná (správna), nevalidná (zavádzajúca), žiadna alebo dvojité (obojsmerná) nápoveda pre ľavé (LVF) aj pravé (RVF) vizuálne pole. Vyhodnocovali sa priemerné RT a pomery medzi nimi (pozri ďalej). Správnosť

detekovania cieľových podnetov slúžila ako vylučovacie kritérium – do analýzy neboli zaradení účastníci s presnosťou nižšou ako 70% pre ktorúkoľvek podmienku. Z analýzy boli tiež vylúčení účastníci s dominantnou ľavou rukou, ako aj účastníci uvádzajúci obojručnosť. Ojedinelé extrémne hodnoty (outliers) boli upravené winsorizáciou.

V tab. 3 sú sumarizované výsledky pacientov a v tab. 4 výsledky zdravých účastníkov. Čísla v zátvorke znamenajú štandardnú odchýlku od priemeru skupiny.

| Pacienti | Priemerný RT (ms) | | Presnosť (%) | |
|------------|-------------------|----------|--------------|--------|
| | LVF | RVF | LVF | RVF |
| Kongruent. | 576(166) | 590(200) | 91(18) | 92(19) |
| Nekongr. | 640(159) | 655(203) | 87(17) | 86(19) |
| Validná | 573(169) | 567(169) | 91(18) | 89(20) |
| Nevalidná | 635(166) | 654(231) | 81(27) | 86(17) |
| Dvojitá | 587(156) | 616(218) | 89(17) | 90(17) |
| Žiadna | 634(165) | 631(141) | 90(19) | 90(22) |

Tab. 3: Priemerné reakčné časy a správnosť vo všetkých podmienkach u pacientov.

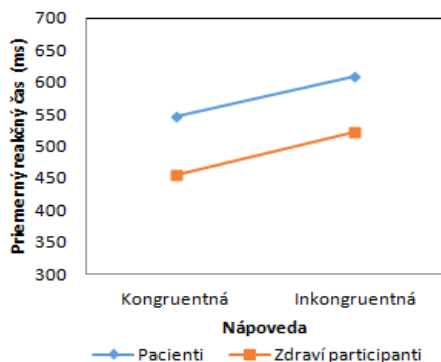
| Zdraví | Priemerný RT (ms) | | Presnosť (%) | |
|------------|-------------------|----------|--------------|-------|
| | LVF | RVF | LVF | RVF |
| Kongruent. | 458 (111) | 447(94) | 98(4) | 98(2) |
| Nekongr. | 518 (121) | 516(112) | 94(10) | 93(7) |
| Validná | 446(100) | 441(99) | 99(2) | 98(3) |
| Nevalidná | 520(130) | 497(108) | 94(9) | 92(8) |
| Dvojitá | 469(123) | 467(95) | 95(7) | 96(6) |
| Žiadna | 517(121) | 519(109) | 96(7) | 98(3) |

Tab. 4: Priemerné reakčné časy a správnosť vo všetkých podmienkach u zdravých účastníkov.

V rámci tohto experimentu sme si položili viacero otázok. Testy sme vyhodnocovali pomocou softvérového balíka SPSS.

A/ Testovali sme efekt kongruentnosti (zhodného a nezodného) smerovania šípok (distrakčných a cieľovej) na reakčný čas pomocou 3-faktorového testu ANOVA pre opakované merania, s medziskupinovým faktorom a vnútrogrupinovými faktormi vizuálne pole a kongruentnosť: dizajn $2 \times 2 \times 2 =$ (zdraví, pacienti) \times (LVF, RVF) \times (kongruentnosť, nekongruentnosť). Ako kofaktor bol zaradený vek účastníkov. Zistili sme štatisticky významný hlavný efekt faktora skupina ($F(1,27) = 6,84, p < 0,05$) a vysoko signifikantný kofaktor vek ($F(1,27) = 13,75, p < 0,005$). Na obr. 6 sú znázornené priemerné reakčné časy skupiny pacientov a kontrolnej skupiny zdravých účastníkov zvlášť pre kongruentnú a nekongruentnú podmienku. Je zrejme,

že skupina pacientov mala dlhšie reakčné časy v porovnaní s kontrolnou skupinou, a to v oboch podmienkach. Kratšie reakčné časy pre kongruentnú podmienku v porovnaní s nekongruentnou neboli štatisticky významné. Podobne, nezistili sme efekt faktora vizuálne pole.

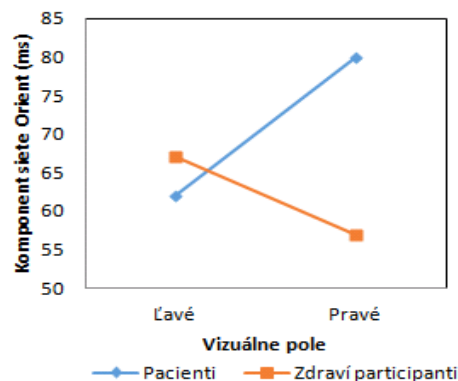


Obr. 6: Priemerné reakčné časy skupiny pacientov a kontrolnej skupiny pre kongruentnú a nekongruentnú podmienku.

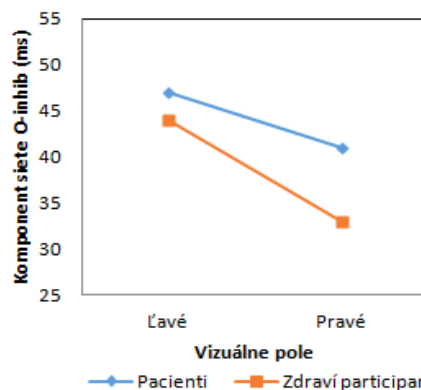
B/ Testovali sme efekt nápovedy na RT pomocou testu ANOVA s dizajnom $2 \times 2 \times 4 =$ (zdraví, pacienti) \times (LVF, RVF) \times (valídna, nevalídna, dvojité, žiadna). Podobne ako pri efekte kongruentnosti šípok, našli sme štatisticky významný efekt faktora skupina ($F(1,27) = 6,78, p < 0,05$) a kofaktora vek ($F(1,27) = 13,75, p < 0,005$). Efekt ostatných faktorov, rovnako ako interakcie medzi nimi, sme nezistili.

C/ Na zistenie rozdielov v pozornosťných sieťach (C, A, O, Oi, Of) sme použili test ANOVA pre opakované merania s dizajnom $2 \times 2 =$ (zdraví, pacienti) \times (LVF, RVF). Našli sme štatisticky významný hlavný efekt faktora vizuálne pole pre sieť O ($F(1,28) = 11,25, p < 0,005$), čo ilustruje obr. 7, a komponent Oi ($F(1,28) = 9,12, p < 0,01$), čo je znázornené na obr. 8. To znamená, že vizuálno-priestorová sieť O a inhibičná vizuálno-priestorová sieť Oi vplývali rozdielne na spracovanie informácie prezentovanej v ľavej a v pravej polovici zorného poľa. Pre komponenty siete O ($F(1,28) = 12,35, p < 0,005$) aj Oi ($F(1,28) = 8,05, p < 0,01$) sme našli aj významnú interakciu faktora vizuálne pole s kofaktorom vek. Teda účastníci spracúvali informácie sprava a zľava rozdielne v závislosti od veku. Štatisticky významná bola aj interakcia skupina*vizuálne pole ($F(1,28) = 4,84, p < 0,05$) pre komponent O. Pacienti a zdraví účastníci teda dokázali využiť valídnu (správnu) nápovedu (v porovnaní s nevalídnu nápovedou) rozdielne podľa toho, v ktorej polovici zorného poľa bol podnet prezentovaný. Ako vidieť na obr. 7, pacienti mali vyšší rozdiel v RT medzi valídnu a nevalídnu informáciou v prípade podnetov prezentovaných v pravej polovici zorného poľa (RVF), v porovnaní s podnetmi prezentovanými naľavo. V kontrolnej skupine bol tento

efekt opačný: výraznejší efekt nápovedy bol pri podnetoch prezentovaných v ľavej časti zorného poľa v porovnaní s podnetmi prezentovanými napravo.

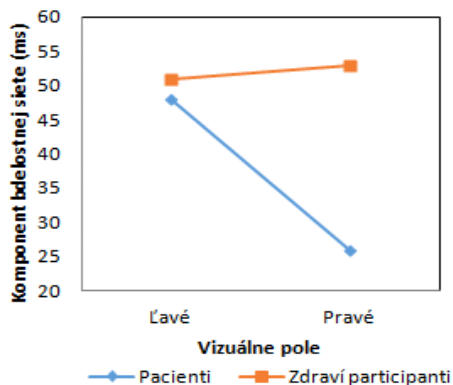


Obr. 7: Porovnanie oboch skupín z pohľadu pozornosťnej siete (O) pre obe vizuálne polia.



Obr. 8: Porovnanie oboch skupín z pohľadu pozornosťnej siete (Oi) pre obe vizuálne polia.

Pre ostatné komponenty pozornosti (C, A, Of) sme nezistili významné efekty žiadneho z faktorov a ani ich interakcie, s výnimkou trendu, ktorý sme pozorovali pre pozornosťnú sieť A (alerting) pri interakcii skupina*vizuálne pole ($F(1,28) = 3,04, p = 0,092$). Znáčornenie poskytuje obr. 9.



Obr. 9: Porovnanie oboch skupín z pohľadu pozornosťnej siete (A) pre obe vizuálne polia.

4 Diskusia

V kognitívnom teste pracovnej pamäti dosiahli zdraví účastníci vyššie skóre v porovnaní so skupinou pacientov po NCMP a to v oboch podmienkach (pri opakovaní sekvencie číslíc v rovnakom i opačnom poradí).

Ide o test numerického charakteru, keďže ako položky boli použité čísllice. Účastníci, ktorí nikdy neprišli do kontaktu s testom podobného typu dosiahnu spravidla nižší výkon už z dôvodu, že nezhlukujú položky. Pokiaľ ide o tých, čo majú nejakú skúsenosť s týmto testom, môžeme konštatovať, že v ich prípade meriame schopnosť zhlukovať. Avšak v prípade reverzného testu numerickej pracovnej pamäti to nemusí celkom platiť, pretože tam je zhlukovanie náročnejšie. Predpokladáme však, že na nižšie skóre u pacientov mohla mať vplyv aj v priemere menšia skúsenosť s daným typom testu. Akú úlohu u pacientov zohrala cievná príhoda, nevieme dobre posúdiť, pretože sme mali k dispozícii len údaje o tom, ktorá hemisféra bola poškodená a stupeň poškodenia. Údaje o presnejšej lokalizácii poškodenia sme nemali k dispozícii.

V použitom teste jemnej motoriky sme nenašli štatisticky významné rozdiely medzi skupinami. Svoju úlohu mohol zohrať aj fakt, že naša skupina pacientov nemala výrazné motorické deficity. Dôležitým faktorom je pravdepodobne aj momentálna psychická pohoda pri vykonávaní testu.

Pozornostný test LANT považujeme za najdôležitejší test nášho výskumu. Silnou stránkou testu je jeho viacnásobná faktorovosť (meranie reakčného času v mnohých podmienkach, navyše z hľadiska lateralít pozornostných sietí). V tomto krátkom článku sme si nemohli klásť za úlohu podať kompletnú správu o našom použití tak komplexného testu, akým je LANT. Našou úlohou bolo skôr poskytnúť ilustráciu o možnostiach metódy, ktorá používa tento test a informovať o momentálnom stave nášho výskumu.

Našli sme štatisticky významné rozdiely v kognitívnom stave pacientov a kontrolnej skupiny týkajúce sa rôznych experimentálnych podmienok a vizuálnych polí. Tieto naše zistenia si zaslúžia aspoň krátky komentár a diskusiu.

Pri štatistickom testovaní rozdielov medzi skupinou pacientov a skupinou zdravých účastníkov v kongruentnej a nekongruentnej podmienke sme zistili, že pacienti dosiahli štatisticky významne dlhšie reakčné časy v oboch podmienkach. Podobný, štatisticky významný medziskupinový rozdiel sme našli aj v reakčných časoch pre podmienku s náповедou validnou, nevalidnou, dvojistou a žiadnou. Celkovo dlhšie reakčné časy pacientov v porovnaní so zdravými dobrovoľníkmi však nemusia byť priamym efektom neurologického poškodenia. K pozorovanému rozdielu mohla prirodzene dopomôcť aj skutočnosť, že pacienti boli testovaní po noci strávenej v spánkovom

laboratóriu (kde boli počas spánku monitorovaní aj pomocou EEG elektród pripojených na povrch hlavy, čo značne znižuje komfort spánku). Pokiaľ ide o zdravých pacientov, k lepšiemu výsledku mohla prispieť aj skutočnosť, že až na dvoch účastníkoch, išlo o ľudí s vysokoškolským vzdelaním (teda vyšším, ako väčšina pacientov), ktorí sa dokázali dobre namotivovať na test.

Pokiaľ ide o jednotlivé komponenty pozornostných sietí, našli sme štatisticky významnú interakciu (skupina*vizuálne pole) v sieťach O (orienting). Pacienti a zdraví jedinci teda vykazovali pre túto sieť rozdielnu lateralitu. Zdraví dobrovoľníci sa viac spoliehali na náповedu pri podnetoch prezentovaných v ľavej časti zorného poľa. U pacientov bola táto lateralizácia efektu náповedy opačná: väčší vplyv náповedy (väčší rozdiel v reakčných časoch) vidíme pri podnetoch prezentovaných v pravom zornom poli. Výsledok zdravých účastníkov je v súlade so súčasnými poznatkami, že práva hemisféra je dominantná pokiaľ ide o vizuálno-priestorovú pozornosť (Rinne, Hassan a Goniotakis, 2013; Green, 2008). Môže to súvisieť s poškodením pravej hemisféry, čo sme osobitne nerozlišovali.

Okrem toho sme zistili aj rozdiel v spracovaní zavádzajúcej informácie vzhľadom na zorné pole, a to bez ohľadu na skupinu (zdravých účastníkov či pacientov). Obe skupiny sa nechali viac zmiatať nevalidnou (zavádzajúcou) náповедou pri podnetoch prezentovaných v ľavom zornom poli. Toto zistenie môže poukazovať na automatickejšie spracovanie podnetov z pravého zorného poľa.

Z ostatných predbežných výsledkov je ešte zaujímavý v medziskupinovom porovnaní odlišný trend v sieťi A (alerting), pokiaľ ide o spracovanie informácií v ľavom a pravom vizuálnom poli. V pravom vizuálnom poli je u pacientov rozdiel medzi reakčnými časmi vo validnej podmienke a v podmienke bez náповedy výrazne nižší. Pacienti teda neboli schopní dostatočne dobre využiť upozornenie na nasledujúci podnet, pokiaľ bol tento prezentovaný v pravej časti zorného poľa. Je možné, že aj tento jav mohol byť spôsobený rôznym stupňom poškodenia jednotlivých hemisfér. Na overenie tejto domnienky je nutné rozšíriť našu vzorku o väčšie množstvo pacientov s lateralizovaným kôrovým poškodením.

Vo všetkých štatistických testoch sme našli vysoko významný kofaktor vek účastníkov. Tento mal na kognitívny výkon rozdielny efekt v skupine pacientov ako v kontrolnej skupine. V ďalšej analýze sa budeme tomuto javu venovať podrobnejšie.

Čo sa týka slabých stránok výskumu:

(a) Zatiaľ neboli štatisticky testované rozdiely medzi pacientmi NCMP a kontrolnou skupinou v presnosti odpovedí. Toto testovanie by mohlo mať tiež relevantnú výpovednú hodnotu.

(b) Mali sme pomerne malý počet pacientov s poškodením ľavej a pravej hemisféry (okrem pacientov NCMP s poškodením mozgového kmeňa

a s cerebrálnym poškodením). Z tohto dôvodu nebola možná osobitá analýza so zameraním na lateralizáciu poškodenia.

5 Záver

Naše výsledky naznačujú špecifické deficity v kognitívnom výkone pacientov s NCMP, najmä v inhibičnej zložke pozornostnej siete O (orienting). Pozorované efekty vykazujú odlišnú lateralizáciu v porovnaní so zdravou populáciou. Na podrobnejšie preskúmanie vplyvu lateralizácie je nutné preveriť efekt lokalizácie CMP. Presné a podrobné preskúmanie behaviorálno-psychologickej úrovne poškodenia kognitívnych funkcií pacientov s NCMP sa javí ako vysoko opodstatnené. Výsledky takéhoto výskumu môžu prispieť k nielen objasneniu ich patofyziológie, no aj k následnej terapii pacientov s NCMP.

PodĎakovanie

Ďakujeme MUDr. Igorovi Riečanskému, PhD. za zapožičanie licencie softvéru E-prime 2.0 potrebnej na vytvorenie počítačovej verzie dvoch z použitých testov. Taktiež ďakujeme Mgr. Matejovi Pecháčovi za prípravu testu LANT a implementáciu pamäťového testu i motorického testu.

Táto štúdia bola podporená projektom MZ SR 2012/56-SAV-6.

Literatúra

Ambler, Z., Bauer, J. (2010). *Cévní onemocnění mozku*. Klinická neurologie, Praha: Triton.

Chen, X. a spol. (2015). Validation of NINDS-CSN neuropsychological battery for vascular cognitive impairment in Chinese stroke patients. *BMC Neurology*, 2015, s. 15–20.

Fan, J., McCandliss, B. D., Thomas, K. M., a Posner, M. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14(3), 340–347.

Gazzaniga, M.S., Ivry, R.B., Mangun, G.R. (2014). *Cognitive Neuroscience. The Biology of the Mind*. W.W. Norton, 4. vyd.

Greene, D. J., Barnea, A., Herzberg, K., Rassis, A., Neta, M., Raz, A., & Zaidel, E. (2008). Measuring attention in the hemispheres: The Lateralized Attention Network Test (LANT). *Brain and Cognition*, 66, 21–31.

Kaufman, A.S.; Lichtenberger, E. (2006). *Assessing Adolescent and Adult Intelligence*. Hoboken (NJ): Wiley, 3. vyd.

Mellon, L. a spol. (2015). Cognitive impairment six months after ischaemic stroke: a profile from the ASPIRE-S study. *BMC Neurology*, s. 15–31.

Moran, G.B. a spol. (2014). Fatigue, psychological and cognitive impairment following transient ischaemic attack and minor stroke: a systematic review. *European Journal of Neurology*, 21, 1258–1267.

Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.

Posner, M., Reichle, M.E. (1994): *Images of Mind*. New York: Scientific American Library.

Posner, M., Fan, J. (2013). Attention following stroke. *Neurology*, 81, 782–783.

Rinne, P., Hassan, M., Goniotakis, D., Chohan, K., Sharma, P., Langdon, D., Soto, D. a Bentley, P. (2013). Triple dissociation of attention networks in stroke according to lesion location. *Neurology*, 81(9), 812-820.

Wolf, T.J., Barbee, A.R., a White, D. (2011). Executive dysfunction immediately after mild stroke. *OTJR. Occupation, Participation and Health*, 31(1 suppl), S23-S29.

Wolf, T.J., Rognstad, M. (2013). Changes in cognition following mild stroke. *Neuropsychological Rehabilitation*, 23(2), 256–266.

„Řízená náhoda“ v metakreativitě

Aleš Svoboda

Univerzita Karlova v Praze, Fakulta humanitních studií
U Kříže 8, 158 00 Praha 5
ales.svoboda@fhs.cuni.cz

Abstrakt

Prověřování možností metakreativity, tedy vytvoření autonomních naprogramovaných strojů schopných produkovat výstupy srovnatelné s lidskými uměleckými artefakty, se v dosavadní historii vždy opíralo o metodu „řízené náhody“. Zdrojem inspirace pro esteticky ocenitelné artefakty se zpočátku stávaly jednak struktury produkované křivkami osciloskopů, jednak aditivně skládané struktury řízené parametry nabývajících náhodných hodnot. V obou případech se vždy shodně zapracovává předem daný rozsah variability, která jistě vymezuje rozsah očekávaných inovací. Je možné rámcově odhadnout očekávatelný rozsah složitosti a zároveň i odvodit další strategie jejího překonání další zapracovanou náhodností? V praktické aplikaci lze prověřit skladbu libovolných křivek sestavených z pohybu složených kružnic (Fourierovy transformace) v poloměrech a úhlové rychlosti přizpůsobených poměrům zlatého řezu, případně v další fázi hledat postupy „školení“ křivek pomocí genetických algoritmů.

1 Cybernetic serendipity a Computer graphic

Rok 1968 je pro počítačové umění poměrně významným milníkem, a to ve vzácné shodě jak pro světovou, tak domácí historii. Britská kurátorka polského původu Jasia Reichardt (*1933) se na počátku šedesátých let v souvislosti s organizováním výstavy konkrétní poezie dostala do kontaktu s německým matematikem, filozofem a zakladatelem informační estetiky Maxem Bensem (*1910). Pravděpodobně především on jí významně rozšířil výhled do nově vznikající umělecké oblasti využívající stejně nově se rodící počítačovou techniku. Max Bense organizoval na Stuttgartském Technickém institutu sérii výstav a přednášek autorů, kteří hledali způsob, jak počítače využít pro posílení uměleckých kreativních možností. V oblasti vizuálního umění k nim patřili hlavně dva Benseho žáci – Frieder Nake (*1938) a Georg Nees (*1926). Reichardtová se také, prostřednictvím různých odborných periodik (např. časopisu *Data Systems* a *Computers and Automation*) seznámila i s Američany Michaelem Nollem a Bélou Juleszem z Laboratoří firmy Bell Telephone, Charlesem

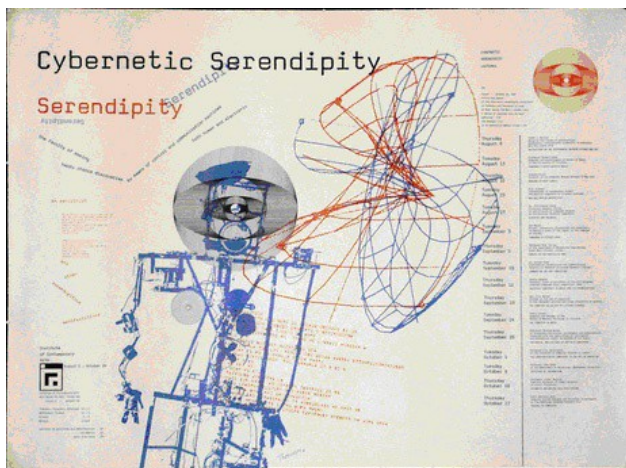
A. Csuriem a s tokijskou skupinou Computer Technique Group (CGT).



Obr. 1. Expozice Cybernetic Serendipity, Institute of Contemporary Arts, Londýn, 1968

Výstava, kterou v Institutu současného umění v Londýně nakonec připravila (od 2. srpna do 20. října), měla název Cybernetic serendipity a po dvojtečce název pokračoval vysvětlujícím podtitulem – Počítač a umění. Ona „serendipita“, pojem vypůjčený od Horace Walpolea, je „šťastnou náhodou“, „nezaslouženou výhodou“, pokladem, který člověku jen tak padne do klína, a pro nadšené zastánce počítačového umění postihoval obzor nově objevených možností. Záběr kurátorské aktivity byl obrovský a pokrýval téměř vše, co se na obzoru počítači podporovaného umění ve světě objevilo: poezie Edwina Morgana *Computer's first Christmas card*, hudba avantgardních hudebníků Johna Cage, Iannise Xenakise a Petera Zinovieffa, robotické realizace Nam June Paika, Edwarda Ihnatowitze, Bruce Lacey a Jamese Seawrighta, filmy Kennetha Knowltona, Michaela Nolla, Nicholase Negroponteho a Johna Whitney st., a samozřejmě celá řada „počítačových grafik“ (Charles Csuri, Frieder Nake, Michael Noll, Computer Technique Group). Ačkoliv bylo ústřední myšlenkou výstavy prověřit úlohu kybernetiky v současném umění, výběr exponátů tento záměr mnohdy svobodně překračoval, takže vystavena byla i tradiční malba s námětem počítače (Lowell Nesbitt) nebo malující

kinetická plastika Jeana Tinguelyho. Nicméně účast zmíněných zakladatelských osobností a velkých průmyslových, obchodních a vědeckých korporací (IBM, Boeing, General Motors, Westinghouse, Calcomp, Bell Telephone Labs a US Airforce Research Labs) dodávala akci, coby manifestaci možností počítačového umění, na okázalosti. Nová oblast tvorby uměleckého znaku a způsobu uměleckého sdělování se dobově také promítala i do zájmů sémiotických výzkumů a účast Umberta Eca (*1932) byla právě důsledkem jeho delšího zájmu o programované a počítačové umění¹.



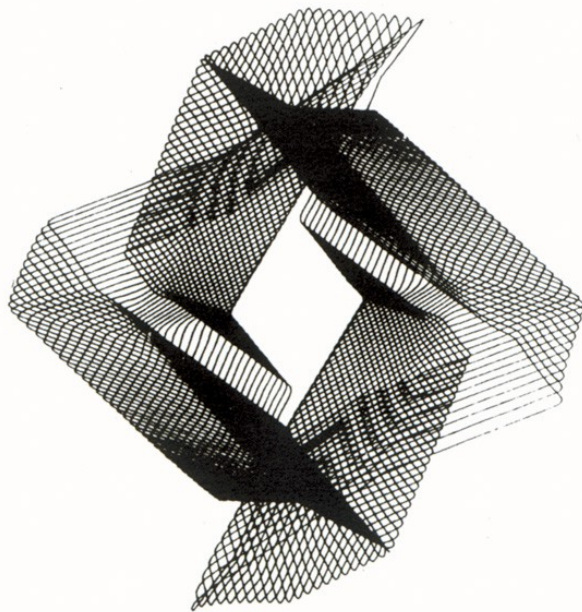
Obr. 2. Cybernetic Serendipity, plakát 1968

Výstava tak v kontextu dějin umění bývá pokládána za první pokus předvést všechny stránky počítačové kreativity (tedy ve výtvarném umění, hudbě, poezii, tanci, plastice, animaci), přičemž jako důležitou součást tvorby byla možné rozpoznat zapojení náhody. Na plakátě příznačně dominují dva vizuální znakové odkazy: dálkově řízený Robot K-456 z roku 1965 Nam June Paika a síťová kresba složených křivek.

Ve stejném roce, dokonce o čtyři měsíce dříve, tedy v dubnu roku 1968, mohli návštěvníci brněnského Domu umění zhlédnout první výstavu počítačového umění ve střední a východní Evropě. Výstava měla ještě reprízu ve Zlíně a její kolektivní charakter byl umožněn osobní sbírkovou činností a mezinárodními kontakty kurátora, mladého teoretika umění, umělce a sběratele Jiřího Valocha (*1946). Podařilo se mu shromáždit a na výstavě představit rozsahem sice skromný, nicméně poměrně

¹ Hiršal, J., Grögerová, B. (1967). Ve svém příspěvku z roku 1965 do časopisu *Nesyo* o programovaném umění mimochodem Eco v úvodu konstatuje, že se v současném umění rozlišují dvě kategorie umělců. Za prvé ti, kdo hledají nové formy a projevují úsilí o „pythagorovský ideál matematické harmonie“. Používají geometrické zákonitosti, ať už euklidovské či ne. Dále ti, kteří „rozpoznali plodnost náhody a ne-rádu“, jsou si vědomi nového vědeckého hodnocení kauzálních a statistických procesů.

reprezentativní výběr aktuálního výtvarného počítačového umění vytisknutého na plotrech – tehdy teprve pět let staré novince. Výstavě dal název *Computer graphic* a kalkovým slovem „komputer“ také v doprovodném textu nový stroj důsledně pojmenovává. Ve výběru byli zahrnuti zásadní autoři ze Spojených států (Charles Csuri, Michael Noll), Kanady (Leslie Mezei) a Německa (Frieder Nake, Georg Nees) a jediný autor tuzemský – Lubomír Sochor.



Obr. 3. Lubomír Sochor: Počítačová grafika, 1968

V úvodním textu *Programované umění* se Valoch (Valoch 1968) snaží zodpovědět otázku po důvodech zařazení počítačových grafik do sféry umění. Konstatuje, že se zúčastnění autoři při práci s počítači začali „zabývat i estetickou stránkou grafických záznamů“ a začali programovat struktury, které „považovali za esteticky cenné“. Zároveň se Valoch snaží o předběžnou kategorizaci programovaných děl. Vymezuje dva druhy vystavených prací = „stochastické grafiky“ a „grafická řešení soustavy rovnic“. Do první kategorie zařazuje práce Nakeho a Neese, v nichž hraje hlavní roli zapojení generátoru náhodných čísel. Autor totiž určuje pouze „pole možností“, ve kterém je počítači ponechána „volba“, výběr konkrétní realizace. Druhou třídu děl reprezentuje právě náš Lubomír Sochor. Grafika je tu přesným grafickým řešením programu zadaného jako soustava rovnic, a to bez možnosti volby. Ve výsledku plotter pouze přeneše vypočítaná data do vizuální podoby. Společně mají obě kategorie to, že usilují o „...vytvoření estetických struktur, minimálně závislých

na lidském subjektu jako tvůrci.“ Začlenění takových artefaktů do domény umění ponechává Valoch na divácích, kteří, pokud jim status umění přisoudí, budou je také specifickým způsobem vnímat a budou tak „realizovat“ jejich estetickou informaci. Podobně jako ve stejné době Bense nebo Eco připomíná Valoch blízkost pojetí „konkrétního umění“, které zaměřením na geometrii, materiál a metodu práce potlačovalo psychologizující tendence dosavadního umění. Taková díla „...zprostředkovávají [...] estetické hodnoty, které jsou primárními hodnotami každého uměleckého díla...“ (Valoch, 1968).

Lubomír Sochor byl od října 1966 zaměstnán v Ústřední vojenské nemocnici jako programátor na analogových počítačích. Podle vlastních slov v autorském příspěvku do katalogu *Computer graphic* (Valoch, 1968) využíval při testování francouzského počítače NADAC 1000, patřícího tehdy do kategorie velkých počítačů, programování grafik ke zjišťování vlastností počítače. Ve zkušebním provozu, při dlouhodobých řešeních se Sochorovi zdály grafické výstupy jako nejvhodnější typ kontrolního testu. Probouzely ovšem zároveň v autorovi zájem o vlastní estetické kvality²

O matematické stránce svých grafik Sochor prozrazuje, že jejich základem byla periodická funkce času $\sin \omega t$ a její derivace $\cos \omega t$. Jde tedy o řešení rovnice

$$y'' + ky' + ay = 0$$

„Prakticky se řeší soustava těchto rovnic s různými koeficienty, jejich řešení a derivace se v různém poměru sčítá. Jeden součet se přivádí na osu x a druhý na osu y souřadnicového zapisovače.“ Sochor v katalogu poznamenává, že „graficky nejzajímavější grafy při součtech:

$$x = b_1y_1 + b_2y_2$$

$$y = b_3y_2 + b_4y_1$$

nebo

$$x = b_1y_1 + b_2y_2' + y_1 (b_5y_3 + b_6y_4')$$

$$y = b_3y_2 + b_4y_1' + y_1 (b_7y_4 + b_8y_3')$$

Změnou koeficientů k, a, b se získá prakticky nekonečné množství různých grafik.“ A dále: „Použití jiné periodické funkce času místo funkce $\sin \omega t$, např. trojúhelníkové, získá se při stejných podmínkách zápisu řada „neméně zajímavých“ grafik. Zajímavé jsou i grafiky, kde se použijí některé průběhy sinusové a některé trojúhelníkové.“ (Valoch, 1968)

Na závěr tohoto stručného historického diskursu snad ještě poznamenejme, že blízko k našemu tématu měly v roce 1968 také dvě newyorské výstavy – *Some more beginnings: an exhibition of submitted works involving technical materials and processes* v Brooklyn Museum

² Na základě získaných zkušeností napsal Lubomír Sochor společně s Milanem Morávkem studii *Estetické hodnoty mimoestetických sdělení*, která vyšla ve *Výtvarné práci* (1967/11).

a výstava s názvem *The machine, as seen at the end of the mechanical age* v Museum of Modern Art. Ovšem kurátorským zaměřením se jako velmi podstatná jeví ještě geograficky blízká výstava ze série záhřebských „Tendencí“ – *Tendencije 4* (Brown, 2008: 76). Pořádala se v městském Centru pro kulturu a informace v srpnu 1968 a byla doprovázena kolokviem Počítače a vizuální výzkum. Na dobových autentických snímcích můžeme rozpoznat Sinusového člověka od Charlese A. Csuriho, snímky z jeho animace Kolibřík, obrazové básně a grafické deformace Leslie Mezeie, síťové kresby Petera Milojeviće – tedy řada realizací založených na programování složených křivek.

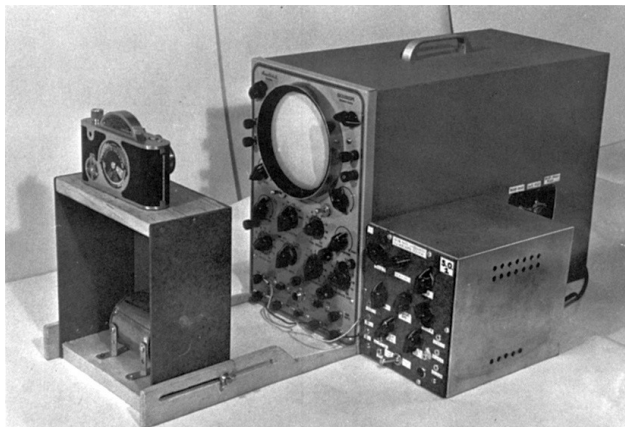


Obr. 4. Pohled do instalace výstavy *Tendencije 4*, 1968

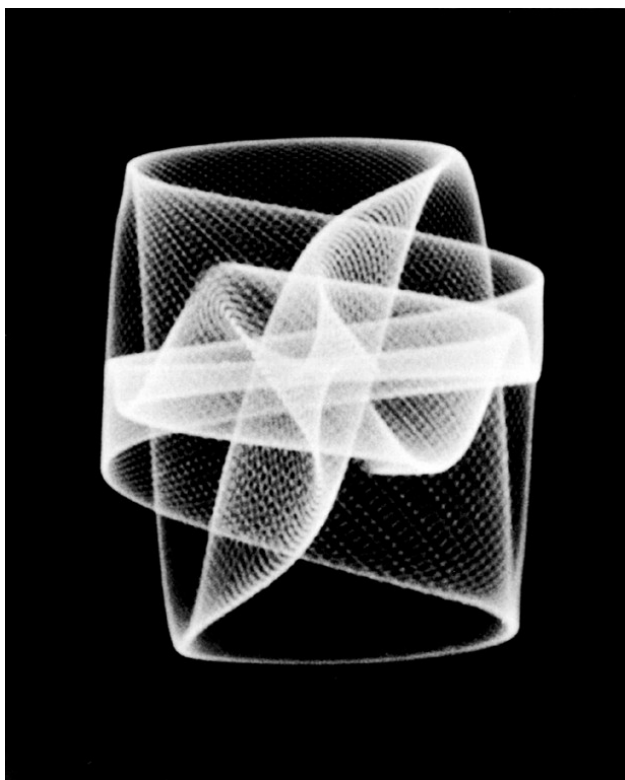
2 Estetika matematických funkcí a modernistických prostorových abstrakcí

Často se připomíná, že digitální počítačové umění má analogové počátky. První výpočetní stroje bývaly konstruovány na analogových základech a stejně tak první umělecké kontakty se uskutečnily právě s nimi. Již od roku 1950 se výtvarník a matematik z Cherokee ve Spojených státech Ben F. (Benjamin Francis) Laposky (*1914) zabýval tvorbou grafik na analogových počítačích. Nazýval je „oscillons“ (oscilóny) a od roku 1953 je pravidelně vystavoval, nejprve jen černobíle, ale od roku 1956 i barevně. Prakticky využíval určitý druh osciloskopu, přičemž výstupy na obrazovce zachycoval standardním fotoaparát s delší expoziční dobou. Periodické funkce vykreslovaly působivé grafické struktury, jejichž pozitivní estetické hodnocení lze zřejmě

přičíst komplementaritě potěšení ze stupňujícího se komplikovaného opakování doprovázeného přehlednými organizujícími symetriemi.



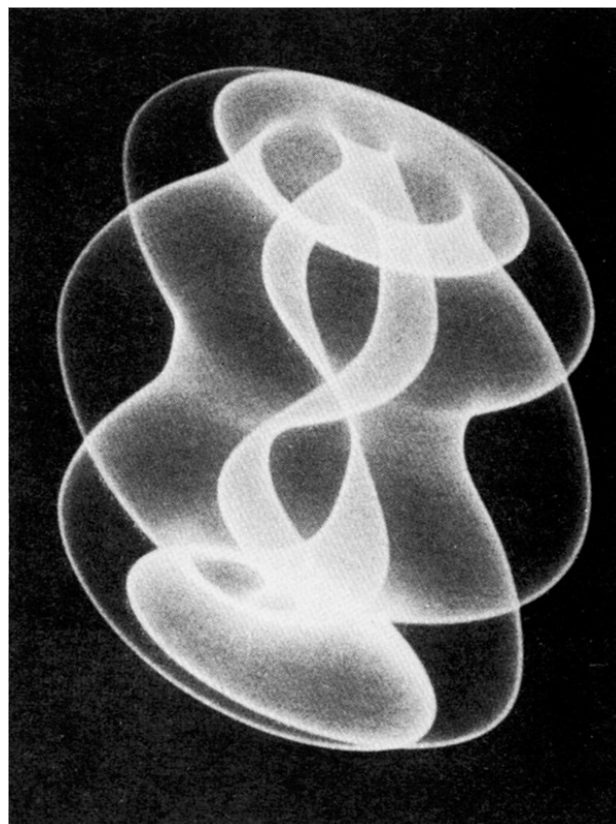
Obr. 5. Ben F. Laposky: Zařízení k pořizování oscilónů



Obr. 6. Ben F. Laposky: Oscilóny, 1953

Velmi brzy a nezávisle na Lapoském začal s analogovými osciloskopy experimentovat i Rakušan Herbert W. Franke (*1927). Vědec, matematik a programátor vykročil na pole počítačové grafiky a digitálního umění. Od poloviny 50. let vytvářel grafiky nazvané

„oszilogramme“ (oscilogramy), které převáděl do podoby sitotiskových grafik.

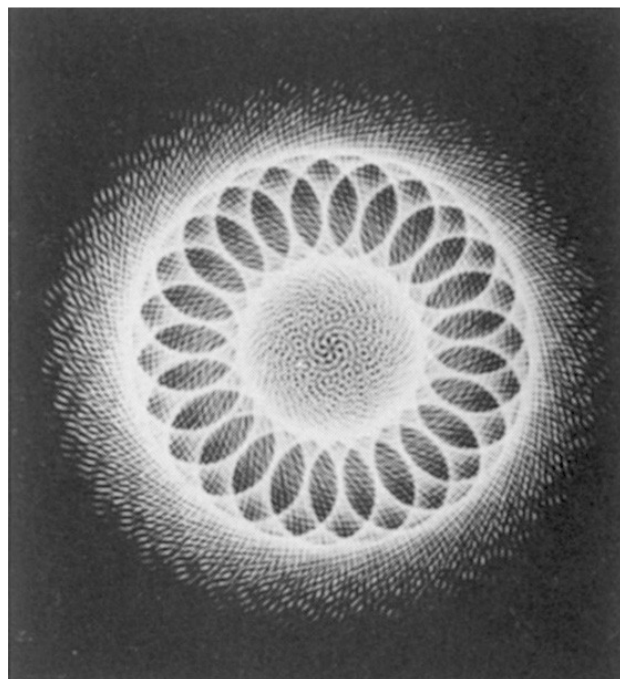


Obr. 7. Herbert W. Franke: Oscilogram, 1956

Na vizuální potěšení z tvarových možností analogových postupů podávaných osciloskopy, na jejich spletité efemérní tvary kreslené světelným paprskem pak Franke v druhé polovině 60. let navázal s novou technologií – plotterem – tvorbou cyklu Algebraických křivek. Jako umělec si však byl vědom i souběžné nabídky počítačů pro kreativní postupy, a sice práce s náhodou alternující různé průběhy tvarových řešení. V řadě svých realizací se zapojením náhody do algoritmů pracoval.

Američana Johna Whitneye st. (*1918) přivedl k zájmu o využití počítačů experimentální film. V roce 1949 již získal uznání na 1. mezinárodní soutěži experimentálního filmu v Belgii. Následně uskutečnil velmi rané pokusy s počítačovými grafickými systémy (IBM), přičemž výstupu na obrazovce jednoduše zachycoval filmovou kamerou. Pro komplexnost dosažených výsledků je považován za jednoho z „otců počítačové grafiky“. V roce 1957 vytvořil barevný 16 mm film o trvání přibližně sedmi minut Yantra a o čtyři roky později film Katalog. Abstraktní kompozice pořádající dynamické přeskupování světelných kroužků a čtverečků, jejich cyklická organizace spoluorganizovaná náhodnými vlivy

slučovala matematickou pravidelnost a organickou spontánnost. Některé sekvence se pak stávaly samostatnými počítačovými grafikami.



Obr. 8. John Whitney st.: Počítačová grafika, 1964



Obr. 9. Naum Gabo: Konstrukce v prostoru, 1937–39

Na formální podobnost řady síťových počítačových struktur s Gabovými plastikami upozornil Mike King: „V klastru Laposky, Franke a Whitney můžeme

zpozorovat, že by důvěrným přítelem všech tří mohl být Naum Gabo, jehož práce z pozdních 30. let předbíhají počítačová zkoumání o 20 let. Zatímco Gabo uskutečňoval své ideje v plastikách, počítačovní umělci používali osciloskopy a účelově vytvořená zařízení, často přenášející jejich myšlenky spíše do filmu, než do tisku nebo do fotografií.“ (King, 2013)

3 Svět cyklického opakování

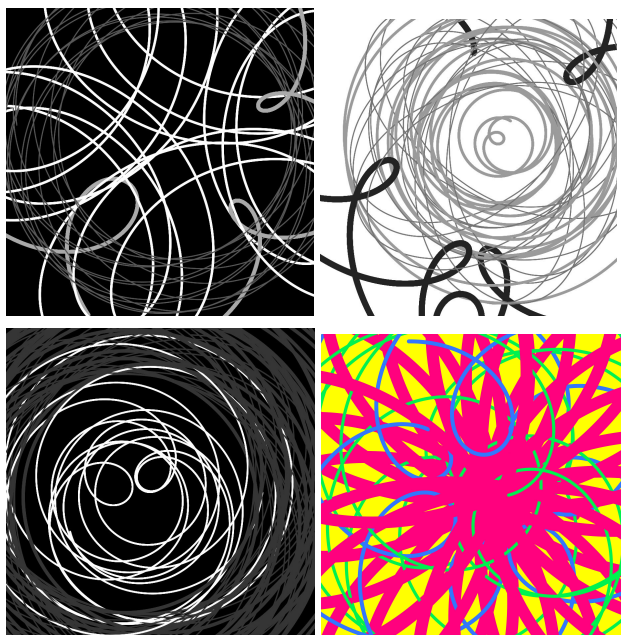
John Barrow, profesor astronomie se zájmem o moderní trendy ve fyzice, astronomii a matematice je polyhistor, který často prokazuje svoji schopnost nalézat spojitost mezi věcným, přírodovědným zkoumáním a estetickým hodnocením. V knize *Vesmírná galerie* (Barrow, 2011) se ve stručných a koncentrovaných kapitolách věnuje vždy určitému obecnému vyznění zcela konkrétního vědeckého tématu. V kapitole *Kadmos a harmonia*, *Siny a kosiny* obrací naši pozornost k významu trigonometrie. Je samozřejmě matematicky prokázáno, že „...téměř libovolnou periodickou křivku lze vyjádřit jako součet souboru sinových a kosinových křivek o různých vlnových délkách.“ (Barrow, 2011: 295) Tato metoda popisu se nazývá aproximací Fourierovými řadami – nekonečný soubor různých sinů a kosinů o různých periodách může dokonale vyjádřit periodické proměny (Barrow, 2011: 296).³ Barrow konstatuje, že sinus a kosinus jsou nejužívanějšími funkcemi v aplikované matematice a příčinu nalézá v tom, že děje, které kolem sebe v přírodě pozorujeme, nastávají tak často proto, že jsou *stabilní*. Nestabilní děje jsou pomíjivé, a proto řídké. Naopak stabilní děje, pokud jsou mírně narušeny, trvají, protože se opět vrací k rovnováze. Velikost výchylky je vždy omezena, tak jak to vyjadřují křivky sinu a kosinu. Jinými slovy, stabilní děje oscilují kolem rovnovážné polohy. Barrow může vyjádřit velmi obecný, srozumitelný a případně vizuálně, názorně postižitelný závěr: „...sinus a kosinus [vlastně] popisují jednoduše a úplně stabilitu světa.“ (Barrow, 2011: 296)

4 Algoritmus oživovaného stereotypu

Spatřovat krásu ve složitých a zároveň v jistém smyslu přehledných vizualizacích matematických funkcí je samozřejmě staršího data, nová je relativní snadnost dosažení tohoto druhu komplexity. Teoretik umění nových médií Lev Manovich ve studii poskytnuté pro katalog výstavy *Abstraction Now* dává velmi podnětně do

³ Tedy funkce $f(x)$ s periodou 2π může být psána jako nekonečná řada $f(x) = a_0 + \sum [a_n \cos(nx) + b_n \sin(nx)]$ pro $n = 1, \dots, \infty$, kde a_0 , a_n a b_n jsou konstanty, které lze snadno určit z funkce $f(x)$, jež je takto vyjádřena pomocí Fourierovy řady.

kontrastu „reduktivní“ fázi děl zakladatelů nezobrazivého umění a komplikované, nicméně určitým zákonům podléhající abstrakce současných neomezených umělců: „... pokud modernistická abstrakce předpokládá, že za senzuační bohatostí světa stojí jednoduché abstraktní struktury, které vytvářejí všichni tuto bohatost, v softwarové abstrakci takové oddělování úrovně chybí. Místo toho poznáváme dynamickou interakci prvků, které opakovaně vedou k jistým zákonitým konfiguracím.“ (Pfaffenbichler a Droschl, 2003: 84) Je pak samozřejmě možné pokračovat k celkem přijatelnému zobecnění: „... jsme svědky systému, který vypadá, že se mění stav od stavu, osciluje mezi řádem a chaosem – znovu, přesně analogicky ke komplexním systémům nalezeným v přirozeném světě.“ (Pfaffenbichler a Droschl, 2003: 84).



Obr. 10. Aleš Svoboda: 4 výstupy z programu Kroužení, 2016

Předložený program (Kroužení) navazuje na průkopnické období počítačového umění a rozpracovává postup, který sestavuje periodickou křivku složenou z náhodně zvoleného počtu kruhových průběhů, které mají náhodný úhlový počátek, náhodný přírůstek a náhodné průměry kružnic. Rozsahy proměnných samozřejmě musí splňovat určité meze, které jednak respektují velikost zadaného formátu, jednak mohou vytvářet vizuálně spojitý průběh vykreslujících bodů. Komutativní charakter skládání jednotlivých průběhů dovozuje zapojit i nulovou velikost průměru či přírůstek, a tím redukovat maximální rozsah skládaných kružnic. I při těchto omezeních dochází k velké variabilitě kombinatorických možností parametrů

kruhů, a tím i k bohatému vzájemnému strukturování. Lze vyzorovat, že důsledky provázání velikostí parametrů mají podobně periodický charakter. Další pozorování by mělo přivést ke stanovení hodnotících kritérií, které by mohly ovlivňovat výběr parametrů při tvorbě genetického algoritmu.

Poděkování

Tento příspěvek vznikl za podpory Programu rozvoje vědních oblastí na Univerzitě Karlově v rámci dílčího programu P18 Fenomenologie a sémiotika pod číslem projektu FHS/P18/2016/16.

Literatura

- Barrow, J. D. (2011). *Vesmírná galerie*. Praha: Argo/ Dokořán, ss. 292–296.
- Brown, P., Gere, Ch., Lambert, N., Mason, C. (ed.). (2008). *White Heat Cold Logic*, British Computer Art 1960–1980, Cambridge, Mass. – London: The MIT Press.
- Hiršal, J., Grögerová, B. (ed.). (1967). *Slovo, písmo, akce, hlas*. K estetice kultury technického věku. Výběr z esejí, manifestů a uměleckých programů druhé poloviny XX. století. Praha: Československý spisovatel.
- King, M. (2013) Computers and Modern Art, in Digital Art Museum. <http://digitalartmuseum.org/essays/king02.htm> [16. 7. 2013]
- Pfaffenbichler, N., Droschl, S. (ed.). (2003). *Abstraction Now* (katalog). Vína: Künstlerhaus Wien, Camera Austria, ss. 80–85)
- Popper, F. (1993). *The art at the electronic age*, New York: Harry N. Abrams.
- Rosen, Margit (ed.). (2010). *Bit International: A little-known story about a movement, a magazine, and the computer's arrival in art: New Tendencies and Bit International, 1961-1973*. ZKM, Center for Art and Media Karlsruhe – Cambridge, Mass.: The MIT Press.
- V&A Pattern – Digital Pioneers*. (2009). Londýn: V&A Publishing.
- Valoch, J. (1968) *Computer graphic*. Brno: Katalogy Domu umění 93/DU.

EEG detekce emoční valence: Frontální alfa asymetrie jako klasifikátor valence emocí

Bc. Michael Tesař, Mgr. Michal Vavrečka, PhD.

Katedra pedagogiky a psychologie, Jihočeská univerzita
Dukelská 9, 370 01 České Budějovice
michtesar@gmail.com, michalvavrecka@gmail.com

Abstrakt

Frontální alfa asymetrie (FAA) vychází z konceptu dvou základních tezí. Lze detekovat afektivní odezvu pomocí EEG v oblasti mPFC. Zároveň se odezvy nejspolehlivěji detekují ve frekvenčním pásmu od 8 do 13 Hz. Řada výzkumů se opírá o analýzy klidových záznamů s aspirací na odhalení markerů v signálu EEG, zejména pak u afektivních poruch. Markantní pozornost je v recentních pracích věnována markerům u depresivních pacientů. Z hlediska nových trendů se FAA používají experimentálně u léčby afektivních poruch formou metody neurofeedback. Analýzou FAA u 24 zkoumaných osob měřených 111 kanálovým EEG, bylo zjištěno, že nelze rozlišovat emoční valenci na základě rozdílné afektivně-vizuální stimulace.

1 Úvod

Spektrální analýza EEG v oblasti frontální alfa (8 - 13 Hz) asymetrie je experimentálně využívána jako biomarker afektivních poruch. Oblast orbitofrontálního kortexu je funkčně považována jako mediátor afektivní odezvy (Quaedflieg, Smulders et al. 2016).

Podstatou analýzy frontálních alfa asymetrií je získání inverzního koeficientu z dvou párových elektrod čelního laloku. Typicky se pro tyto účely používají elektrody Fp1-Fp2 nebo AF3-AF4. Pro vypočtení koeficientu frontálních alfa asymetrií se zpravidla užívá rozdíl přirozeného logaritmu pravého výkonu ve spektru od levého výkonu ve spektru.

Relativně zvýšená FAA odráží vyšší aktivaci levého čelního laloku a naopak její pokles odkazuje na zvýšenou pravostrannou aktivitu. Z recentních nálezů

se pravostranné FAA podílí na vzniku depresivních a nostalgických nálad (Tullett, Wildschut et al. 2015). Naopak levostanná FAA naznačuje slibné výsledky v copingu u stresových reakcí (Dusing, Tops et al. 2016).

Davidson (1993) definoval klidovou frontální alfa asymetrii jako relativně stabilní klasifikátor motivační a emoční složky člověka. Je ověřeno, že tento jev je stabilní od narození do raného dětství (Howarth, Fettig et al. 2016).

Quaedflieg, Smulders et al. (2016) potvrzují, že FAA je jev měřitelný metodou neurofeedback a lze výzkumné subjekty trénovat ke snižování koeficientu FAA. Pro tyto účely využívají tzn. individuální alfa asymetrii tak, že hodnoty adjustují pro každý výzkumný subjekt zvlášť.

Cílem práce je ověřit, zda je možné využít frontální alfa asymetrie u evokovaného signálu sérií afektivně elicitujících vizuálních podnětů a ty následně rozlišit v signálu EEG podle emoční valence prezentovaného podnětu.

2 Metody

2.1 Výzkumný vzorek

Výzkumný vzorek čítá po vyloučení některých měření z důvodu kontroly kvality signálu 23 subjektů (12 žen, 11 mužů) o průměrném věku 24,45 (min = 19, max =

37) let. Ve vzorku se též vyskytovaly dva leváci. Teoretická východiska ovšem deklarují, že frontální alfa asymetrie nejsou závislé na leváctví či práváctví. U všech subjektů byl kladen zejména dvě kritéria a sice, aby nebyly pod psychoaktivní medikací a neměly v anamnéze psychiatrickou poruchu. Dále bylo nutné, aby měly standardní vizuální percepci, přičemž přípustná byla pouze korekce zraků brýlemi, které měly subjekty při experimentu nasazeny.

2.2 Experimentální úloha

Experiment se sestává ze systematické vizuální stimulace s emočním nábojem, za současného snímání EEG. Jako afektivně elicitující podněty jsou zvoleny fotografie z databáze IAPS (International affective picture stimuli).

Z nich je vybráno 22 fotografií pro 3 třídy, dle jejich emoční valence na podněty pozitivní, negativní a neutrální. Jejich expozice je právě 3000 ms. Mezi podmětový interval je variabilní úsek mezi 1000 a 1500 ms, pro zamezení anticipace stimulu subjektem.

3 EEG analýza

3.1 Předzpracování dat

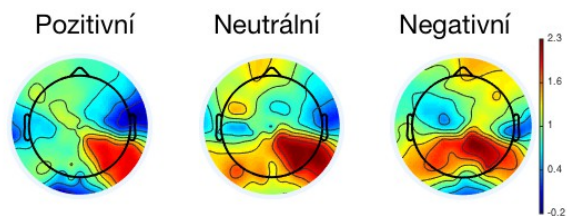
Elektroencefalografický záznam je proveden pasivními Ag-Cl 111 kanálovým EEG o vzorkovací frekvenci 1024 Hz. Referenční elektroda byla zvolena standardně na Nz a následně je re-referencována na průměr všech 111 kanálů. Data jsou po jejich vizuální inspekcí a manuálního vyčištění od artefaktů převzorkována na nižší vzorkovací frekvenci 250 Hz. Během čištění je vyjmuto několik kanálů, které jsou však následně interpolovány (průměrně jde o 3,54 interpolovaných kanálů na subjekt).

V průměru je manuálním čištěním dat odstraněno 19% dat, což reprezentuje 11 epoch. Následně je zjištěno, že

rozložení epoch pro jednotlivé typy stimulů jsou rovnoměrně rozloženy se středem okolo 18 epoch.

3.2 Fourierova transformace signálu

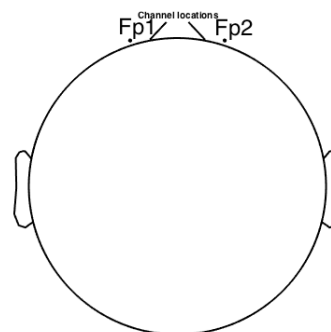
Dále je provedena Fourierova transformace dat na matici 125 datových bodů frekvence a 625 datových bodů času. Rozlišení je nastavené po 1 Hz, pomocí metody Hanning window s 90% wavelet oknem. Výsledná distribuce spekter je prezentována topografickým modelem viz Obr. 1.



Obr. 1: Distribuce signálu po spektrální analýze topografického modelu v čase 2600 ms, kde je vizuálně zaznamenaný jev ve frekvenčním pásmu 8 – 13 Hz. Statisticky je jev na elektrodách Fp1 a Fp2 nevýznamný.

3.3 Výpočet FAA

Takto připravená data pro všechny 3 typy afektivní odezvy jsou podrobena spočtení koeficientu frontálních alfa asymetrií po každý daný časový bod. Vzniká tím vektor logaritmického rozdílu dvou bipolárních elektrod v prefrontální oblasti, konkrétně Fp1 a Fp2 viz Obr. 2.

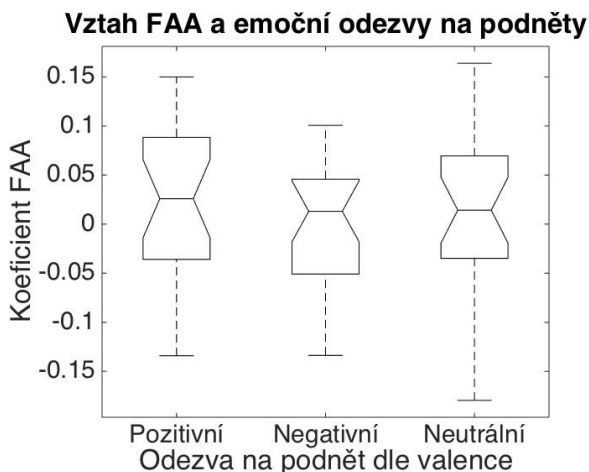


Obr. 2: Lokalizace použitých frontální elektrody Fp1 a Fp2 pro výpočet FAA

4 Výsledky

Data byla nejprve podrobena testování na distribuci dat vzhledem k normálnímu rozložení Shapiro-Wilkovým testem pro každou emoční třídu zvlášť. Kde rozložení koeficientu frontální alfa asymetrie pro pozitivní podněty je $p = 0,12$, negativní podněty $p = 0,23$ a neutrální podněty $p = 0,90$. Lze prohlásit, že na hladině významnosti alfa $p < 0,01$ je distribuce dat rozložena neparametricky.

Vzhledem k tomu byla data následně testována neparametrickými ekvivalenty statistických metod a sice Kruskal-Wallisovým testem analýzy rozptylu. Test na hladině významnosti alfa $p < 0,01$ prokázal, že neexistuje statistický rozdíl ($H(2) = 0,63$, $p = 0,73$) mezi skupinami jednotlivých odezev na vizuálně elicitující podněty. Výsledky jsou v Obr. 3 přehledně vizualizovány.



Obr. 3: Vizualizace výsledků Kruskal-Wallisova testu analýzy rozptylu. Data nejsou signifikantně odlišná.

5 Závěr

Z výsledků je zřejmé, že metodou v tomto nastavení nelze detekovat emoce pomocí FAA.

Na výsledky by mohlo mít vliv, že data nebyla vyčištěna analýzou nezávislých komponent (ICA), neboť se jedná o kritickou oblast prefrontálního kortexu, kde jsou okulární artefakty manifestní. Pokud by FAA měly v metodě neurofeedback fungovat jako

klasifikátor emoční valence vizuálních podnětů je nutné navrhnout jiný typ analýzy.

Samotné výsledky mohou být ovlivněny relativně malým výzkumným vzorkem nebo krátkou sadou podnětů. Pro další výzkum autoři navrhuji využít většího vzorku a současně zvětšit baterii vizuálních stimulů. V této souvislosti je nutné zmínit, že databáze IAPS je sice standardizovaná, ale na velmi malém vzorku subjektů a lze tedy pokládat její vnitřní validitu za nespolehlivou. Z tohoto důvodu autoři také navrhuji využít jiné podnětové databáze.

Z dostupných zdrojů jsou FAA používány pro detekci afektivních poruch. Experimentálně se na kontinuálním záznamu objevují publikaci implikující možnosti FAA jako klasifikátoru nálady, tedy relativně dlouhodobého nastavení emoční odezvy na obecné podněty. Navzdory těmto výsledkům se autoři domnívají, že lze FAA použít jako klasifikátor emoční valence, nicméně je nutné navrhnout zcela nové paradigma testování.

Literatura

- Dusing, R., M. Tops, E. L. Radtke, J. Kuhl and M. Quirin (2016). "Relative frontal brain asymmetry and cortisol release after social stress: The role of action orientation." *Biol Psychol* 115: 86-93.
- Howarth, G. Z., N. B. Fettig, T. W. Curby and M. A. Bell (2016). "Frontal Electroencephalogram Asymmetry and Temperament Across Infancy and Early Childhood: An Exploration of Stability and Bidirectional Relations." *Child Dev* 87(2): 465-476.
- Quaedflieg, C. W., F. T. Smulders, T. Meyer, F. Peeters, H. Merckelbach and T. Smeets (2016). "The validity of individual frontal alpha asymmetry EEG neurofeedback." *Soc Cogn Affect Neurosci* 11(1): 33-43.
- Tullett, A. M., T. Wildschut, C. Sedikides and M. Inzlicht (2015). "Right-frontal cortical asymmetry predicts increased proneness to nostalgia." *Psychophysiology* 52(8): 990-996.

Na cestě k uměle inteligentním systémům: Vyhodnocování inteligence umělých systémů

Ondřej Vadinský

Katedra informačního a znalostního inženýrství VŠE v Praze
Náměstí Winstona Churchilla 4, 130 67 Praha 3, ČR
E-mail: ondrej.vadinsky@vse.cz

Abstrakt

Tento příspěvek se zabývá otázkou, jak poznat a vyhodnotit, zda je umělý systém inteligentní. Sleduje proto uchopení pojmu inteligence ve filosofii a kognitivních vědách. Nejprve vymezuje umělou inteligenci (UI) pomocí distinkcí *silná–slabá*, resp. *obecná–specifická*. Dále se věnuje *Turingovu testu* a jeho rozšířením. Všímá si také provázanosti s kognitivními schopnostmi, jak zdůrazňuje proud *psychometrické UI*. Vyvrcholením snahy uchopit pojem inteligence je Legg-Hutterova *definice univerzální inteligence* a z ní odvozený Legg-Venessův *test algoritmického IQ*. Závěrem příspěvek shrnuje možnosti vyhodnocování inteligence umělých systémů založené na zmíněné definici a testu a zamýšlí se nad možnostmi jejich rozpracování do podoby ucelené metodiky.

1 Úvod

Otázka, jak poznat a vyhodnotit, zda je umělý systém inteligentní, stála již u zrodu umělé inteligence jako vědecké disciplíny. Tato otázka úzce souvisí s definicí a chápáním inteligence či myšlení. Pro vývoj v oboru umělé inteligence je dostatečně přesná definice klíčová tím, že umožňuje tento vývoj směřovat, a nejde tak o pouhé tápání ve tmě, byť i to dokáže přinést zajímavé poznatky. Tento původní směr však byl postupně upozaděn zaměřením se na řešení specifických problémů a dílčích úloh. Až v nedávné době se snaha inteligenci přesněji definovat ideálně včetně nalezení způsobů, jak inteligenci systému vyhodnocovat, opět stává předmětem podrobnějšího výzkumu a to zejména v dílčím oboru označovaném jako *obecná umělá inteligence*.

V sekci 2 příspěvek vymezí různá chápání hlavního cíle disciplíny umělé inteligence. Linii *Turingova testu*, jeho východisek a snah o jeho rozšíření bude sledovat sekce 3. Vztahu inteligence a kognitivních schopností se bude věnovat sekce 4. Přístup definující inteligenci a následně odvozující z této definice prakticky realizovatelný test popíše sekce 5. Konečně sekce 6 shrne možnosti, jak těchto přístupů využít při testování inteligence umělých systémů.

2 Vymezení disciplíny umělá inteligence

Hlavní cíl disciplíny umělá inteligence lze chápat různě. Z filozofie přichází rozlišení mezi *silnou* a *slabou UI* (2.1), v samotné komunitě umělé inteligence se poněkud později objevilo nyní preferované rozlišení na *obecnou* a *specifickou umělou inteligenci* (2.2).

2.1 Vymezení silná–slabá umělá inteligence

Searle (1980) rozlišuje umělou inteligenci podle toho, zda si za cíl klade vysvětlení myslí (*silná UI*), nebo jde pouze o nástroj pro modelování a simulování myslí (*slabá UI*). *Silnou UI* pak identifikuje zejména s tradičním symbolickým přístupem k umělé inteligenci, který je tak chápán jako vysvětlení fungování myslí.

Tuto distinkci však lze interpretovat poněkud volněji, pokud se zaměříme na její normativní aspekt namísto aspektu deskriptivního. Za *silnou UI* lze považovat takový systém, který by byl srovnatelně mocný jako lidská inteligence – zejména co do její univerzálnosti, a přítomnosti dalších kvalit, které s inteligencí u lidí spojujeme, např. rozumění. Oproti tomu *slabá UI* je systém, který by srovnatelně univerzálnosti nedosahoval, a šlo by pouze o užitečný nástroj pro řešení konkrétních problémů.

2.2 Vymezení obecná–specifická UI

Výše zmíněná interpretace Searla je blízká poměrně nedávnému rozlišení *specifická–obecná umělá inteligence*, jak ho vymezuje např. Goertzel (2014). *Specifická UI* nabízí systémy schopné řešit jednu či několik úzce vymezených úloh, které mohou být jednoduché, ale také značně sofistikované. Mimo tuto svou množinu úloh však *specifická UI* selže. *Obecná UI* pak usiluje o vytvoření systémů, které jsou schopné obecného řešení problémů, resp. obecného inteligentního jednání. Na základě toho pak dokáží řešit širokou množinu úkolů ať již jednoduchých, nebo složitých. S *obecnou umělou inteligencí* bývá také spojovaná schopnost přenášet znalosti z řešení jedné úlohy do řešení jiné úlohy, se kterou se systém dosud nesetkal.

3 Turingův test a jeho rozšíření

Pro umělou inteligenci je klíčovou otázkou: „Co je to myšlení?“ Její uchopení lze najít už u Descarta (3.1). Známejší je však Turingovo zpracování, pro něž se ujalo označení *Turingův test* (3.2). Ač se oblast umělé inteligence ubírala spíše jiným směrem, přesto lze najít přístupy, které Turingovu myšlenku dále rozpracovávají: *Úplný Turingův test* (3.3) a *Skutečně úplný Turingův test* (3.4).

3.1 Descartovo rozlišení stroje a člověka

Descartes (1637) věnuje jednu pasáž své *Rozpravy o metodě rozlišení stroje a člověka*. Všimá si zde dvou vlastností, které podle něj náleží pouze člověku: univerzálnost myšlení a schopnost rozumné řeči.

Univerzálností myšlení rozumí Descartes (1637) všestrannost rozumu, tedy schopnost řešit různé problémy, byť ne nutně nejlépe. Rozumnou řeč pak chápe jako schopnost odpovídat na položené otázky.

3.2 Turingův test

Turing (1950) si otázky, co je myšlení, všimá v podobě „Mohou stoje myslet?“, avšak při jejím uchopení provádí následující obrat: Ptá se, zda stroje mohou napodobit lidské chování a myšlení do takové míry, kdy běžný člověk nedokáže rozlišit, zda rozmlouvá s člověkem, či se strojem. Pro toto pojetí se později vžil název *Turingův test*. Descartem uvedené vlastnosti jsou v něm implicitně přítomny, schopnost rozumné řeči je pak pro úspěch stroje v testu Turingem považována za klíčovou.

Turing (1950) založil svůj test na imitační hře, ze které odstranil osobní kontakt hráčů, aby jediné, podle čeho může rozhodčí určit, kdo je člověk a kdo stroj, byly odpovědi na jeho otázky přenášené přes terminál. Obsah otázek není omezen a lze tak jejich prostřednictvím zkoumat i jiné než jen jazykové schopnosti stroje. Zde se tedy může projevit ona zmiňovaná univerzálnost myšlení.

3.3 Úplný Turingův test

Původní *Turingův test* však má své problémy. Protože komunikace mezi strojem a člověkem probíhá přes terminál, tedy v čistě jazykové doméně, je upozaděn vztah jazyka ke světu, tedy jak výpovědi reagující na dění ve světě, tak i výpovědi způsobující dění ve světě. Tohoto si ostatně všimá i Searle (1980), když upozorňuje na problematiku rozumění jazykovým výpovědím.

Kromě toho *Turingův test* ponechává stranou (na jazykovém zprostředkování) i další projevy lidské inteligence v široké paletě lidského chování, které běžně označujeme za inteligentní. Proto Harnad (1991) navrhuje tzv. *úplný Turingův test*, kdy testovaným subjektem

je robot interagující s lidmi v reálném světě. Předmětem testu tak jsou jak jazykové schopnosti ve vztahu ke světu, tak i ostatní lidské inteligentní chování. Lze tedy důkladně prověřit, zda robot jazyku rozumí, např. tím, že zvládne vykonat úkol podle jazykových instrukcí, ale i popsat řeči dění ve světě. Descartem uváděnou všestrannost myšlení lze zkoumat i přímo tak, jak se ukazuje v inteligentním jednání.

3.4 Skutečně úplný Turingův test

I takto rozšířený test má však své limity. Ty vyniknou na pozadí filozofické teorie externalizmu, která upozorňuje na to, že význam jazykových promluv není zcela uvnitř mysli jednotlivce, ale jde o sdílenou kolektivní záležitost v rámci jazykové komunity. Jazykové schopnosti a inteligentní chování tak mají druhový rozměr, jak upozorňuje Schweizer (2012).

Schweizer (2012) proto navrhuje tzv. *skutečně úplný Turingův test*, kdy není testován jeden konkrétní subjekt, ale celý druh či kognitivní typ ve své komunitě. Cílem je zjistit, zda si tento kognitivní typ dokáže vyvinout své vlastní inteligentní chování a jazyk. K zodpovězení otázky, zda stroje (či jiné kognitivní typy) mohou myslet, tak nestačí, pokud si stroj osvojí inteligentní chování a jazyk již existující, a tedy pouze parazituje na chování a jazyku jiného druhu (lidí).

4 Intelligence a kognitivní schopnosti

Intelligence však není osamocená vlastnost, ale souvisí s dalšími kognitivními schopnostmi (4.1). V tomto ohledu je inspirativní oblast tvorby *kognitivních architektur* (4.2). Úzkou vazbu na lidské kognitivní schopnosti pak má *psychometrická umělá intelligence* (4.3).

4.1 Kognitivní paradigma

Za ústřední myšlenku kognitivního přístupu považuje de Mey (1992) nutnost existence modelu světa (de Meyem označovaného jako *náhled na svět*) v systému zpracovávajícím informace.

Tento model, či modely, pak mají prostřednictvím formulovaných očekávání dopad na vnímání, které je jimi vždy zprostředkované, a je tedy výsledkem interakce subjektu a objektu. Konkrétní podobu těchto interakcí de Mey (1992) popisuje pomocí *stratifikovaného modelu vnímání*. V tomto modelu přistupuje subjekt k objektu po částečně nezávislých vrstvách a jednotlivé rysy objektu mapuje na pojmové struktury ve svém modelu světa, čímž vzniká výsledný amalgám vjemu.

Vnímání spolu s jednáním následně umožňuje získávání znalostí. Těm de Mey (1992) přisuzuje zejména podobu možných interakcí subjektu s objektem. Proces, ve kterém se formují lze označit jako *explicitňování implicitního*. Jak subjekt interaguje s ob-

jektem učí se nejprve implicitně struktury možných interakcí. Pokud subjekt v interakcích pokračuje, tyto implicitní struktury mohou nabýt explicitní podoby. Od té chvíle pak mohou řídit další interakce. Inteligence je takto zasazena do komplexu kognitivních schopností.

4.2 Kognitivní architektury

Podobná je i motivace tvorby *kognitivních architektur*. *Kognitivní architektury* jsou doménově nespecifické, tedy obecné, výpočetní modely zachycující esenciální struktury a procesy v mysli. Svou doménovou nespecifičností stojí v protikladu expertních systémů. Tím, že jsou to modely existujících kognitivních systémů, však nejde o umělé inteligence v pravém slova smyslu, ale spíše nástroje k poznání mysli. (Sun, 2007)

Sun (2007) pro jejich tvorbu navrhuje použít *integrováný hierarchický přístup*. V tomto pojetí různé vědy zkoumají kognici na různých úrovních abstrakce: společenské, psychologické, komponentové a fyziologické. Tím kladou na návrh architektur různá omezení jak zdola, tak i shora. Integrováním jednotlivých poznatků různých věd tak lze zaručit vysokou kognitivní realističnost vytvářené architektury.

4.3 Psychometrická umělá inteligence

Jiný přístup čerpá z psychologie, konkrétně z jejího odvětví psychometrie. Ta se zabývá systematickým měřením psychologických vlastností (zejména inteligence) u lidí pomocí různých testů. Bringsjord a Schimanski (2003) navrhuje, že by toho UI měla využít a být pojatá jako *psychometrická*. Cílem by tak mělo být vytvářet systémy, které budou dosahovat solidních výsledků ve všech zavedených a validovaných testech inteligence a dalších mentálních schopností.

Je však otázkou, nakolik je možné bez úprav převzít testy postavené na psychologických definicích lidské inteligence. Jejich antropomorfnost může být na škodu. Zřejmě bude potřeba tyto testy vylepšit a zobecnit, jak navrhuje Besold a spol. (2015).

5 Definování inteligence

Dosud představené přístupy neodpovídají na otázku: „Co je inteligence?“ příliš konkrétně a již vůbec ne formálně. V nedávné době se však na poli *obecné umělé inteligence* objevily nové snahy směřující k explicitnímu a formálnímu zodpovězení této otázky. Výsledkem je zejména *definice univerzální inteligence* (5.1), ale třeba i pokus o její rozšíření do podoby *definice pragmatické obecné inteligence* (5.2). Přínos těchto snah však spočívá i v možnosti odvodit z formální definice prakticky realizovatelný test jako je *test algoritmického IQ* (5.3).

5.1 Definice univerzální inteligence

Ve snaze o vytvoření formální a precizně formulované definice inteligence, Legg a Hutter (2007) prostudovali řadu definic, teorií a testů zabývajících se inteligencí u lidí, zvířat i strojů. Z jejich společných rysů odvodili následující neformální definici: „Inteligence měří schopnost agenta dosahovat cílů v mnoha různých prostředích.“ Formalizaci této definice pojmenovanou *univerzální inteligence* ukazuje rovnice 1:

$$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_{\mu}^{\pi} \quad (1)$$

kde *univerzální inteligence* Υ agenta π je dána jeho schopností dosahovat cílů popsanou hodnotovou funkcí V_{μ}^{π} jako očekávanou sumu všech budoucích odměn nad množinou prostředí μ vážených pomocí Kolmogorovské složitosti K .

V Legg-Hutterově formální definici stojí za povšimnutí následující prvky (Legg a Hutter, 2007):

- Definice uvažuje prostředí μ , agenta π a jejich po krocích postupující interakci skrze akce agenta a_i , jeho pozorování (vnímání) o_i a odměny r_i pocházející z prostředí. Prostředí je popsáno spočetnou pravděpodobnostní mírou μ vjemů a odměn podmíněných dosavadní historií interakcí.
- Definice zvažuje všechna spočetná prostředí a existenci mnoha hypotéz agenta o aktuálním prostředí. Jako Occamova břitva $2^{-K(\mu)}$ zde slouží Kolmogorovská složitost. Pokud lze pro popis pravděpodobnostní míry prostředí (tj. hypotézu o prostředí) použít krátký program, pak má prostředí malou Kolmogorovskou složitost (ta je založena na délce nejkratšího programu popisujícího sekvenci bitů) a skrze Occamovu břitvu velkou váhu. Komplexní prostředí (hypotézy) mají tedy menší dopad na celkový výkon agenta, než prostředí méně složitá. Ovšem i komplexní prostředí (hypotézy) jsou zvažovány, pokud jsou konzistentní s dosavadní historií interakcí.
- Schopnost agenta dosahovat cíle je popsána hodnotovou funkcí: $V_{\mu}^{\pi} := E(\sum_{i=1}^{\infty} r_i) \leq 1$ v podstatě jako maximalizace očekávaných budoucích odměn r_i na základě minulých interakcí s prostředím. Časová preference je přítomna ve způsobu, jakým prostředí přiděluje odměny. Tedy prostředí (či úloha) samo určuje, zda je lepší pomalé ale přesnější, nebo naopak rychlé ale méně přesné řešení.

Legg a Hutter (2007) považují představenou definici za klíčový prvek, který umožní umělé inteligenci dosáhnout svého cíle, tj. vytvořit skutečně inteligentní systém. Z tohoto hlediska jsou užitečné některé její vlastnosti, zejména schopnost řadit různé agenty podle jejich inteligence, od agentů, kteří se chovají

náhodně, až po agenty chovající se optimálně. Vzhledem k tomu, že definice uvažuje všechna Turingovsky spočetná prostředí, musí být agent pro dosažení vysoké *univerzální inteligence* skutečně obecně inteligentní. Definice staví na fundamentálních pojmech jako výpočet, informace a složitost a není tak kulturně předpojatá nebo antropocentrická. Tato definice by se tak mohla stát východiskem pro vyhodnocování inteligence umělých agentů a položit teoretický základ praktickým testům. Zde je však nutno upozornit, že transformace definice postavené na nespočetné Kolmogorovské složitosti, uvažující všechna spočetná prostředí a odhad agenta týkající se všech budoucích odměn, do podoby prakticky proveditelného testu nebude snadná a nutně si vyžádá nějaký aproximační přístup.

5.2 Definice pragmatické obecné inteligence

Goertzel (2010) uvádí, že má Legg-Hutterova definice několik nedostatků z hlediska aplikace na reálné agenty v reálných prostředích. Goertzelova kritika se soustřeďuje do tří oblastí: zacházení s cíli, zacházení s prostředím a neuvažování výpočetních nároků v definici univerzální inteligence.

Z hlediska způsobu práce s cíli si Goertzel (2010) všímá problematičnosti explicitních cílů definovaných v prostředí a s tím související odměny přicházející z prostředí. Zjevně existují cíle, které si agenti dávají sami, a jen sami agenti vědí, zda je splnili. Navíc zřejmě ne všechno inteligentní chování sleduje nějaké cíle. Goertzelovým řešením je jednak rozšíření definice inteligence o další pravděpodobnostní strukturu popisující cíle, a dále také pohled na tuto definici jako na nepřímou aplikovatelnou na reálné agenty z pozice hypotetických situací. Pro vyhodnocení jejich inteligence je tedy podle Goertzela třeba, aby jiný agent z pozorování chování testovaného agenta vyvodil jeho inteligenci podle upravených měř.

Co se týče prostředí Goertzel (2010) upozorňuje na to, že inteligentní agenti mohou být na některá prostředí adaptovaní, ať už evolučně jako jsou lidé, nebo návrhem jako jsou umělé systémy. Pro vyhodnocení inteligence tak nemusí být vždy zajímavá absolutní obecnost (univerzálnost) agenta ale spíše obecnost, která je vůči některým prostředím předpojatá. Namísto fixace Solomonoff-Levinovy univerzální pravděpodobnostní distribuce prostředí v původní definici, tak Goertzel navrhuje uvažovat různé pravděpodobnostní distribuce prostředí. Pro porovnání s lidmi pak může být zejména zajímavá taková pravděpodobnostní distribuce, která odpovídá rozložení prostředí, ve kterých se lidé běžně vyskytují a inteligentně v nich jednají.

Původní definice neuvažuje žádné aspekty výpočetní náročnosti inteligence, což jak uvádí Goertzel (2010) nereflexkuje to, že se reální inteligentní agenti musí vypořádat s omezenými zdroji. Goertzel proto navrhuje uvažovat jednak omezenou paměť

agentů přímo v definici, a dále zakomponovat do definice možnost normalizace výsledné inteligence podle pravděpodobnostní distribuce spotřebovaných výpočetních prostředků.

Na základě této kritiky Goertzel (2010) zavádí *definici pragmatické obecné inteligence*, jak popisuje rovnice 2:

$$\Pi(\pi) \equiv \sum_{\mu \in E, g \in G, T} \nu(\mu) \gamma(g, \mu) V_{\mu, g, T}^{\pi} \quad (2)$$

kde *pragmatická obecná inteligence* Π agenta π je dána jako jeho schopnost dosahovat komplexních cílů v komplexních prostředích, jak popisuje hodnotová funkce $V_{\mu, g, T}^{\pi}$ jako očekávanou sumu budoucích odměn relativně k pravděpodobnostní distribuci ν prostředí μ a pravděpodobnostní distribuci γ cílů g v časovém intervalu T .

Goertzel (2010) zároveň uvádí mírně modifikovanou hodnotovou funkci $V_{\mu, g, T}^{\pi} \equiv E \left(\sum_{i=s}^t r_g(I_{g, s, i}) \right)$, ve které vztahuje úspěšnost agenta k cíli g zadanému na začátku časového intervalu T , v němž agent bere do úvahy očekávané s cílem spjaté odměny r_g ve všech interakčních sekvencích $I_{g, s, i}$ vybraným podle μ .

Aby zohlednil výpočetní nároky inteligence, zavádí Goertzel (2010) *definici efektivní pragmatické obecné inteligence*, jak popisuje rovnice 3:

$$\Pi_{\text{Eff}}(\pi) \equiv \sum_{\mu \in E, g \in G, Q, T} \frac{\nu(\mu) \gamma(g, \mu) \eta_{\pi, \mu, g, T}(Q)}{Q} V_{\mu, g, T}^{\pi} \quad (3)$$

kde $\eta_{\pi, \mu, g, T}$ udává pravděpodobnost, že agent π v prostředí μ při plnění cíle g v časovém intervalu T spotřebuje výpočetní zdroje Q . Použití pravděpodobnostní distribuce η umožňuje uvažovat nedeterministické agenty, Q je pak pro zjednodušení amalgámem časových, paměťových, energetických a jiných zdrojů a lze ho očekávat někde mezi kladnými reálnými čísly.

Protože pravděpodobnostní distribuce uvažované v Goertzelem upravených definicích nemusejí být Solomonoff-Levinova univerzální pravděpodobnostní distribuce, uvedené sumy nemusí konvergovat a zjišťování podmínek, za kterých k tomu dojde je obtížné. (Goertzel, 2010)

Goertzel (2010) také uvádí počáteční pokus o vymezení obecnosti versus specifčnosti inteligence, který označuje jako *intelektuální šíři agenta*. Jeho přístup spočívá v možnosti vytvořit fuzzy množinu kontextů, vůči kterým je nějaký agent inteligentní. Za kontext Goertzel označuje trojici prostředí, cíl a časový interval. Tuto fuzzy množinu pak lze normalizovat do podoby pravděpodobnostní distribuce a zjistit její entropii. Sám autor však uvádí, že tento přístup je limitovaný tím, že neuvažuje vzájemné závislosti mezi prostředím, resp. cíli, vůči kterým jsou agenti inteligentní.

5.3 Test algoritmického IQ

Legg a Veness (2013a) zkoumali potenciál *definice univerzální inteligence* pro odvození prakticky proveditelného testu. Ve výsledku navrhli *test algoritmického IQ (AIQ test)* aproximující míru *univerzální inteligence* agenta. Transformace původní definice si vyžádala mnoho kroků, až dospěla do podoby zachycené v rovnici 4:

$$\hat{\Upsilon}(\pi) := \frac{1}{N} \sum_{i=1}^N \hat{V}_{p_i}^{\pi} \quad (4)$$

kde *AIQ odhad univerzální inteligence* $\hat{\Upsilon}$ agenta π je dán jeho schopností dosahovat cílů popsanou empirickou hodnotovou funkcí $\hat{V}_{p_i}^{\pi}$ jako celková odměna dosažená agentem za jeden průběh programu prostředí p_i zprůměrováno přes N programů v konečném vzorku programů prostředí.

Z porovnání s definicí *univerzální inteligence* vystávají následující rozdíly (Legg a Veness, 2013a):

- Test uvažuje konečný vzorek N programů prostředí p_i , agenta π a jejich interakci. Stejně prostředí může být popsáno několika programy a stejný program může být zahrnut ve vzorku vícekrát.
- Protože se využívá pouze N programů prostředí, počítá se prostý průměr. Idea Occamovy břitvy je ale zachována ve způsobu, kterým se programy prostředí vybírají do vzorku, a to díky využití Solomonoffovy univerzální distribuce: $M_U(x) := \sum_{p: \mathcal{U}(p)=x} 2^{-l(p)}$. Kratší program tak má vyšší pravděpodobnost zahrnutí do vzorku, i když jsou uvažovány všechny programy popisující dané prostředí a ne pouze ten nejkratší, jak tomu je při použití Kolmogorovské složitosti v Legg-Hutterově definici.
- Důvodem použití empirické hodnotové funkce $\hat{V}_{p_i}^{\pi}$ je testování pouze omezeného počtu interakcí mezi agentem a prostředím. Odměny přidělené prostředím nejsou shora omezené 1 jako u původní definice, a nejsou nijak diskontovány pro vyjádření časové preference na této úrovni. Použije se tedy celková navracená odměna dosažená za daný konečný počet interakcí mezi agentem a prostředím.

Legg-Venessův *test algoritmického IQ* je dostupný jako Open Source prototypová implementace v Pythonu z (Legg a Veness, 2013b). Ta umožňuje testovat agenty dodané v podobě interní implementace nebo v podobě externího programu, který komunikuje s testem prostřednictvím specifického rozhraní. Test zahrnuje jednoduché agenty $freq$, Q_0 , Q_λ a HLQ_λ a rozhraní pro testování externí implementace agenta *MC-AIXI*. Test lze do určité míry konfigurovat, zejména je možné nastavit počet programů prostředí pro odhadování *AIQ skóre* a počet interakcí mezi agentem a

prostředím (délku epizody). Kromě toho lze měnit velikost prostoru pozorování a akcí. Test je postavený na jednoduchém Turingovu stroji (modifikovaný BF referenční stroj), na jehož výběru jsou výsledky testu závislé.

6 Možnosti vyhodnocování inteligence umělých systémů

Na základně výše představeného se otevírá řada cest pro další výzkumnou práci. *AIQ test* lze použít k vyhodnocování inteligence umělých systémů ale i paradigmat umělé inteligence (6.1). Lze zkoumat možnosti, jakými tento test dále rozšířit (6.1). A konečně lze také uvažovat nad vytvořením ucelené metodiky pro vyhodnocování inteligence umělých systémů (6.3).

6.1 Vyhodnocování systémů a paradigmat umělé inteligence

Jeden možný přístup využívající *definici univerzální inteligence* a *test algoritmického IQ* k vyhodnocování systémů a paradigmat umělé inteligence představil Vadinský (2015). Přístup spočívá v provádění formálních analýz systémů umělé inteligence za využití Legg-Hutterovy definice a v testování těchto systémů pomocí *AIQ testu*. Podle rozsahu provedení pak lze v různé míře abstrahovat poznatky o jednotlivých systémech, paradigmatech a do určité míry i o některých ústředních otázkách disciplíny umělé inteligence. Následuje stručně shrnutí:

1. Vyhodnocení předpokladů inteligence konkrétního systému:

- (a) *AIQ test systému* – implementovaný systém lze testovat, přičemž lze očekávat technické obtíže při propojování systému s testem a obtíže z hlediska vysokých výpočetních nároků samotného testu. Výsledky lze využít jednak ke srovnání s jinými již testovanými systémy, ale také ke srovnání s upravenými verzemi téhož systému.
- (b) *Formální analýza systému* – formálně popsaný systém lze analyzovat, přičemž lze očekávat obtíže při hledání vhodného formalizmu popisujícího interakce systému s prostředím a zachycujícího vnitřní strukturu systému. Analýza může ovšem být pojata i méně důsledně jako úvaha, v jak složitých prostředích si systém povede dobře, jak je schopen pracovat s historií a predikcemi interakcí s prostředím. Podle důslednosti analýzy pak lze získat různě přesnou představu o uspořádání systémů podle jejich univerzální inteligence.

2. Vyhodnocení předpokladů inteligence konkrétního paradigmatu:

- (a) *Indukce z výsledků AIQ testů a formálních analýz* – pokud se podaří otestovat a formálně analyzovat více systémů reprezentujících nějaké paradigma umělé inteligence, lze z jejich výsledků indukci dospět k závěrům o daném paradigmatu. Přitom je potřeba vypořádat se s problémy indukce týkajícími se množství testovaných systémů i toho, jak jsou tyto systémy dobře implementované.
- (b) *Dedukce z formálního popisu paradigmatu* – pokud by se podařilo vytvořit formální popis reprezentující určité paradigma umělé inteligence, mělo by z něj být možné pomocí *definice univerzální inteligence* dedukovat nějaké závěry týkající se *univerzální inteligence* systémů vycházejících z toho paradigmatu.

3. Vyhodnocení předpokladů inteligence více paradigmat:

- (a) *Indukce z výsledků AIQ testů a formálních analýz* – výsledky získané v předchozím kroku začnou být zajímavé zejména, pokud by je bylo možné porovnat s výsledky jiného paradigmatu. Poté by mělo být možné validovat hypotézy považující jedno paradigma za lepší než jiné, nebo naopak všechna paradigma za stejně vhodná (tj. samotnou ústřední myšlenku funkcionalismu).
- (b) *Dedukce z formálních popisů paradigmat* – podobně i výsledky dedukcí budou zajímavé pokud půjdou porovnat s dedukcemi o ostatních paradigmatech.

6.2 Rozšíření testu algoritmického IQ

AIQ test vychází z původní *definice univerzální inteligence*. Na základě rozvinutí definice, které podává Goertzel (2010), se nabízí několik možných cest rozvoje tohoto testu. Na zvažování je, zda by podobně jako odvození *AIQ testu* z *definice univerzální inteligence*, šlo odvodit praktické testy pro *definice (efektivní) pragmatické obecné inteligence*. Zřejmě největší výzvou zde bude to, že sumy, na kterých jsou Goertzelovy definice založeny, nemusí konvergovat. Pokud by se toto ukázalo jako nerealizovatelné, mohlo by být zajímavé alespoň modifikovat *AIQ test* o některé Goertzelovy koncepty, zejména aspekt vážení inteligence pomocí využitých výpočetních zdrojů, a také ideu *intelektuální šíře agenta*.

Další možnosti rozšíření vycházejí z experimentů s *AIQ testem*, jak provedl Vadinský (v přípravě). Při testech agenta *MC-AIXI* se ukázalo, že doba trvání jedné interakce agenta s prostředím nemusí být konstantní, ale

může se s narůstajícím počtem již proběhlých interakcí zkracovat. Toto zjištění je samo o sobě nepřekvapivé – odráží určitý způsob zlepšování agenta s narůstajícími zkušenostmi – ale může zachycovat určitý aspekt inteligence, který samotným *AIQ skórem* zachycen není. Podobný aspekt může být vidět v tom, jak rychle, či naopak pomalu výsledky agenta konvergují s růstem počtu interakcí. Zdá se tedy, že při vyhodnocování výsledků *AIQ testu* je potřeba přihlídnout nejen k samotnému *AIQ skóre*. Další postřehy jsou spíše technického rázu. Vzhledem k velkým výpočetním nárokům testu by bylo užitečné umožnit pokračování v testu po předčasném ukončení, ke kterému může dojít například z důvodu vypršení přiděleného času ve výpočetním clusteru. Podobně by bylo možné zefektivnit využití výpočetních prostředků, pokud by šlo kombinovat výsledky z více běhů testu s méně vzorky programů prostředí a dosáhnout tak přesnějšího odhadu *AIQ* bez nutnosti pustit test znovu s více vzorky.

6.3 Metodika vyhodnocování inteligence umělých systémů

Na základě výše prezentovaného lze uvažovat o vytvoření ucelené metodiky pro vyhodnocování inteligence umělých systémů. Tuto metodiku lze budovat zejména kolem *definice univerzální inteligence* a *testu algoritmického IQ*. Na základě reflexe Goertzelovy kritiky se však zdá, že by měly být zahrnuty i jiné dimenze. Těmi jsou zejména *efektivita využití výpočetních zdrojů* a *míra intelektuální šíře agenta*. Zajímavou může být i bližší analýza typů prostředí, ve kterých je agent úspěšný zejména z důvodu možného porovnání s lidskou inteligencí. Otázka, jaké další testy do metodiky zahrnout, však zůstává otevřená a vyžaduje další průzkum.

Další úvahy by se měly soustředit na konkrétní zapojení jednotlivých testů. Pro účely tohoto příspěvku se omezím na některé aspekty využití *AIQ testu*. Při počátečních experimentech s *AIQ testem*, jak popsal Vadinský (v přípravě), vystoupil do popředí zejména problém vysokých výpočetních nároků v kombinaci s velkým konfiguračním prostorem agentů. Některé agenty je možné konfigurovat skrze různé parametry, které ovlivňují jejich výkon, přičemž tyto parametry mohou mít řadu hodnot. Pro komplexní otestování agenta by bylo potřeba prozkoumat co nejvíce těchto kombinací. *AIQ test* odhaduje *univerzální inteligenci* agenta na základě jeho chování v určitém vzorku programů prostředí. Z větších vzorků lze získat přesnější odhad, který je pak lépe porovnatelný s jinými výsledky, ale doba testu se prodlužuje. Pokud se budou pro vyhodnocení agenta používat zejména extrémní hodnoty, např. maxima, lze použít přístup založený na *vícekolovém postupném zpřesňování*. V prvním kole se otestuje co nejvíce konfigurací agenta na relativně malém vzorku programů prostředí. To umožní relativně

| Agent | Konfigurace | | AIQ skóre pro počet vzorků | | | | | |
|---------|-------------|-----------------|----------------------------|------|-------|------|--------|------|
| | # | Parametry | 1 000 | | 3 000 | | 10 000 | |
| MC-AIXI | 1 | (50;8;1;0,05) | 39,2 | ±1,1 | 39,3 | ±0,7 | 39,2 | ±0,4 |
| MC-AIXI | 2 | (50;8;2;0,05) | 34,6 | ±1,1 | | | | |
| MC-AIXI | 3 | (50;8;3;0,05) | 35,0 | ±1,0 | | | | |
| MC-AIXI | 4 | (50;16;1;0,05) | 32,7 | ±1,1 | | | | |
| MC-AIXI | 5 | (50;16;2;0,05) | 37,7 | ±1,1 | 37,9 | ±0,7 | | |
| MC-AIXI | 6 | (50;16;3;0,05) | 37,4 | ±1,1 | 38,0 | ±0,7 | | |
| MC-AIXI | 7 | (100;8;1;0,05) | 38,4 | ±1,0 | 39,0 | ±0,7 | 39,1 | ±0,4 |
| MC-AIXI | 8 | (100;8;2;0,05) | 37,0 | ±1,0 | | | | |
| MC-AIXI | 9 | (100;8;3;0,05) | 34,2 | ±1,0 | | | | |
| MC-AIXI | 10 | (100;16;1;0,05) | 32,7 | ±1,0 | | | | |
| MC-AIXI | 11 | (100;16;2;0,05) | 38,4 | ±1,1 | 38,3 | ±0,7 | 39,1 | ±0,4 |
| MC-AIXI | 12 | (100;16;3;0,05) | 36,3 | ±1,0 | | | | |

Tab. 1: Metoda vícekolového postupného zpřesňování odhadu *AIQ skóre* zvyšováním počtu testovaných vzorků pro 1 000 interakcí agenta s prostředím na referenčním stroji BF 5. Přesnost odhadu je vyjádřena polovinou intervalu spolehlivosti na 5% hladině významnosti.

rychle prohledat velký konfigurační prostor a vytipovat zajímavé konfigurace, které postoupí do dalšího kola. V druhém kole proběhne test s větším vzorkem programů prostředí, aby došlo k přesnějšímu vymezení daných zajímavých hodnot. V závislosti na tom, jak přesné odhady jsou potřeba, kolik je konfigurací a jaké je jejich rozložení, lze celý testovací proces provést ve dvou, nebo více kolech. Limitujícím faktorem je zde to, že současná verze *AIQ testu* neumí kombinovat výsledky z jednotlivých průběhů pro zpřesnění odhadu a při přílišném rozmělnění do mnoha kol by mohl celkový výpočetní čas přesáhnout jednokolový test.

Příklad zachycující prezentovanou myšlenku je ukázán v tabulce 1. Pro test byla použita Monte Carlo aproximace agenta *AIXI (MC-AIXI)* od Veness a spol. (2011). Test tvořilo 1 000 interakcí agenta s prostředím na referenčním stroji BF 5 s prostorem pěti symbolů pro akce, jednoho symbolu pro pozorování a jednoho pro odměnu. V prvním kole bylo testováno 12 konfigurací na vzorku 1 000 prostředí z celkových 200 000 programů prostředí dostupných pro test. Do druhého kola, které proběhlo na vzorku 3 000 prostředí, postoupilo 5 konfigurací. Třetí kolo na vzorku 10 000 prostředí absolvovaly 3 konfigurace. Celkově tak test proběhl na 57 000 vzorcích oproti 120 000, pokud by všechny konfigurace prošly rovnou plným testem. Tři nejlepší konfigurace se však testovaly na 14 000 vzorcích namísto 10 000, protože test neumí kombinovat dílčí výsledky.

7 Závěr

Příspěvek sledoval filozofické a kognitivní předpoklady inteligence ve snaze vyjasnit, k čemu disciplína umělé inteligence směřuje. Pro přesnější uchopení hlavního cíle umělé inteligence posloužily zejména distinkce

silná–slabá a *obecná–specifická umělá inteligence*. Další výklad pak probíhal z pozice *obecné*, resp. *silné umělé inteligence*.

První sledovaná linie se týkala *Turingova testu*, jehož myšlenkové základy lze vysledovat již u Descarta. *Turingův test* samotný vyvolal řadu kritiky, co do své schopnosti zachytit, zda stroje mohou myslet, což vedlo ke snahám o jeho rozšíření. V příspěvku jsou prezentovány dvě: Harnadův *úplný Turingův test* a Schweizerův *skutečně úplný Turingův test*.

Druhou sledovanou linií byly vazby inteligence na ostatní kognitivní schopnosti. Jako východisko zde bylo prezentováno *kognitivní paradigma*. Jako ilustrace přístupů vycházejících z této pozice pak posloužily *kognitivní architektury*. Za blízkou pozici pak byla označena i *psychometrická umělá inteligence*, která vychází z testování lidské inteligence dalších kognitivních schopností.

Třetí sledovaná linie se zaměřila na nedávné snahy o přesné definování inteligence. Popsána zde byla zejména Legg-Hutterova *definice univerzální inteligence*, která abstrahuje a formalizuje definici inteligence z řady definic a testů inteligence z oblastí zabývajících se inteligencí lidí, zvířat a strojů. Jako kritika tohoto přístupu byla ukázána odvozená Goertzelova definice *pragmatické obecné inteligence* a s ní související koncepty. Mezi nimi byl zvláště zdůrazněn koncept *efektivnosti využití zdrojů* a koncept *intelektuální šíře agenta*. Kromě definic jsou pro rozvoj umělé inteligence důležité i prakticky proveditelné testy. Z toho důvodu byl příspěvku představen Legg-Venessův *test algoritického IQ*, který vychází z původní Legg-Hutterovy definice.

Ve své poslední části příspěvek načrtl několik cest dalšího výzkumu. Byly uvedeny úvahy o testování a analyzování systémů a paradigmat umělé inteligence

využívající *AIQ test* a *definici univerzální inteligence*. Tato cesta by měla umožnit srovnání jednotlivých systémů, ale i paradigmat na jejichž základě byly tyto systémy vytvořeny. Při dovedení do nejvyšší úrovně abstrakce by pak tato cesta mohla osvětlit i některé ústřední hypotézy umělé inteligence jako samotný funkcionalismus a podobně. Dále bylo naznačeno několik možností rozšíření *AIQ testu*, kde nejdůležitější z nich je zahrnutí některých konceptů Goertzelovy kritiky *definice univerzální inteligence*. Konečně byly také naznačeny možnosti vývoje ucelené metodiky vyhodnocování inteligence umělých systémů. Tato metodika by měla vycházet z *definice univerzální inteligence a testu algoritmického IQ*, ale jako další dimenze by měla zahrnout Goertzelem představené koncepty *efektivity a intelektuální šíře*. Zde uvedené úvahy jsou však ve svých počátcích a zbývá zde mnoho prostoru pro další práci.

Poděkování

Výsledky prezentované v tabulce 1 byly spočítány s využitím zdrojů Národní gridové infrastruktury MetaCentrum poskytnuté programem „Projekty velkých infrastruktur pro VaVaI“ (E-infrastruktura CESNET LM2015042), za jejichž existenci a zpřístupnění děkuji.

Reference

- Besold, T., Hernández-Orallo, J. a Schmid, U. (2015). Can machine intelligence be measured in the same way as human intelligence? *KI-Künstliche Intelligenz*, str. 1–7.
- Bringsjord, S. a Schimanski, B. (2003). What is artificial intelligence? Psychometric AI as an answer. V *Proceedings of the 18th international joint conference on artificial intelligence (IJCAI'03)*, str. 887–893. IJCAI.
- Descartes, R. (1637). *Rozprava o metodě*. Svoboda, Praha, 3. vyd., 1992.
- Goertzel, B. (2010). Toward a formal characterization of real-world general intelligence V *Proceedings of the 3rd Conference on Artificial General Intelligence, AGI 2010*, str. 19–24.
- Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1):1–48.
- Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1(1):43–54.
- Legg, S. a Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, 17(4):391–444.

- Legg, S. a Veness, J. (2013a). An approximation of the universal intelligence measure. V *Algorithmic Probability and Friends. Bayesian Prediction and Artificial Intelligence*, str. 236–249. Springer.
- Legg, S. a Veness, J. (2013b). AIQ: Algorithmic intelligence quotient [source codes]. Dostupné z WWW: <<https://github.com/mathemajician/AIQ>>.
- de Mey, M. (1992). *The cognitive paradigm*. University of Chicago Press, Chicago and London.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3):417–457.
- Schweizer, P. (2012). The externalist foundations of a truly total Turing test. *Minds and Machines*, 22(3):191–212.
- Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(2):159–193.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- Veness, J., Ng, K. S., Hutter, M., Uther, W. a Silver, D. (2011). A Monte-Carlo AIXI approximation. *Journal of Artificial Intelligence Research*, 40(1):95–142.
- Vadinský, O. (2015). Towards an artificially intelligent system: Possibilities of general evaluation of hybrid paradigm. Besold, T. R., Lamb, L. C., Icard, T. a Miikkulainen, R. (ed.), V *Proceedings of the 10th International Workshop on Neural-Symbolic Learning and Reasoning NeSy'15*, str. 23–29, Buenos Aires. IJCAI.
- Vadinský, O. (v přípravě). Towards an artificially intelligent system: Reproducing AIQ Test Results.

Znalostí meze (super)inteligentních systémů

Jiří Wiedermann

Ústav informatiky AV ČR

Pod Vodárenskou věží 2, 182 07 Praha, Česká republika

Email: jiri.wiedermann@cs.cas.cz

Abstrakt

V příspěvku ukážeme nový pohled na inteligenci založený na znalostním přístupu k výpočtům. Výpočty budeme chápat jako procesy, které generují znalosti nad danou znalostní doménou v rámci příslušné znalostní teorie. V tomto kontextu budeme budeme uvažovat inteligenci jako schopnost získávat informace a transformovat je na znalosti, které jsou dále využívány pro řešení problémů. Hlavním výsledkem příspěvku je poznatek, že pokud je znalostní doména konečná a neměnná, pak lze konstruovat inteligentní systémy s tzv. samo-zlepšující se znalostní teorií, které dříve nebo později dosáhnou takový stav poznání o dané doméně, který již nelze dále kvalitativně vylepšovat. Systém tak dosáhne meze své inteligence.

1 Úvod

Snahy o pochopení rozsahu a mezí lidského vědění patří mezi základní filozofické otázky. Je zřejmé, že jakákoliv omezení, týkající se lidského vědění, musí záviset na smyslových schopnostech lidí, na jejich usuzovacích dovednostech a na jejich chápání pojmu "vědění", resp. "znalosti". Jisté je, že filozofové doposud nebyli schopni dát jasnou odpověď na otázky týkající se hranic lidského poznání. Je tomu tak pravděpodobně proto, že v průběhu věku se naše poznávací a usuzovací schopnosti neustále vyvíjí. To se děje zaváděním různých stále výkonnějších pozorovacích prostředků, prohlubováním znalostí o kognitivních procesech, rozvojem umělé inteligence, epistemologie, apod. Jelikož filozofové se snad až na výjimky z poslední doby nikdy neuchylují k modelování, kterým by zafixovali nějaký stav poznání v dané oblasti, a pracují s implicitními předpoklady, které vyplývají zejména z doby, ve které žijí, je zřejmé, že v jejich úvahách na toto téma se lze jenom těžce orientovat.

Inteligenci lze definovat mnoha způsoby. Pohříchu, z většiny z nich nelze odvodit žádné indície stran možných omezení inteligence. Vezměme např. slovníkovou definici inteligence převzatou z Wikipedie: inteligence je schopnost logického, abstraktního myšlení, porozumění, sebe-uvědomění, komunikace, učení, projevat emoce a rozumět jím, schopnost pamatování, plánování, tvořivosti a řešení problémů. Co pojem, to obrovský filozofický problém, protože není zřejmé, co dané slovo vlastně znamená. Navíc, tato definice je zřetelně antropomorfní, nehodí se např. na zvířata či systémy umělé inteligence. Pro naše účely je proto výhodnější obecnější definice,

kde inteligenci chápeme jako schopnost získávat informace a transformovat je na znalosti, které jsou dále využívány pro řešení problémů anebo pro účely adaptivního chování. Počet "problematických" pojmů se tak sníží na jeden: co je znalost. Pro naše účely budeme chápat znalost jako výsledek specifického procesu, jehož vlastnosti zpřesníme v dalším.

Superinteligenci pak definujeme jako inteligenci, která daleko převyšuje schopnosti lidské inteligence ve všech oblastech, včetně vědecké tvořivosti, všeobecných znalostí a sociálních dovedností [1]. Nutnou podmínkou inteligence jsou tedy znalosti — jejich zaměření na nějakou oblast poznání, jejich rozsah a hloubka, organizace a metody získávání či odvozování nových znalostí. Pokud se tedy zajímáme o nějaké "horní meze" inteligence, je vhodné se (alespoň zpočátku) soustředit na příbuznou otázku: existují nějaká omezení na množství a kvalitu znalostí, které může inteligentní znalostní systém získat o dané znalostní doméně? V příspěvku ukážeme, že na takto formulovaný problém již dovedeme dát smysluplnou, zdůvodněnou odpověď. Východiskem pro nás bude tzv. znalostní teorie výpočtů, která vychází z prací autorů van Leeuwen a Wiedermann (viz např. [6], [7], [8], a [5]).

Dle této teorie jsou výpočty chápány jako procesy, které generují znalosti nad danou znalostní doménou v rámci příslušné znalostní teorie. Hlavním výsledkem příspěvku je poznatek, že pokud je znalostní doména konečná a neměnná, pak lze konstruovat inteligentní systémy s tzv. samo-zlepšující se znalostní teorií, které dříve nebo později dosáhnou takový stav poznání o dané doméně, který již nelze dále kvalitativně vylepšovat. Systém tak dosáhne meze své inteligence. Toto poznání nečekaným způsobem podporuje úvahy J. Horgana o konci vědy [2], které jsou vědeckou komunitou přijímány s rozpaky. Náš výpočetní přístup k inteligenci přináší nové poznatky o mechanismech rozvoje inteligence a zpřesňuje popperovský [4] a kuhnovský [3] pohled na rozvoj vědeckého poznání. Současně vyvrací anebo přinejmenším uvádí na pravou míru některé obecně přijímané představy současných teoretiků umělé inteligence o superinteligenci, která se bude neomezeně rozvíjet pomocí "rekurzivně se zlepšujícího software". Naše modelování ukazuje, že myšlenka samo-zlepšujícího se software neřeší jádro problému: abychom zvýšili intelektuální potenciál inteligentních systémů musíme průběžně zvyšovat rozsah jejich znalostí a korigovat jejich případné "vadné" znalosti, a ne vylepšovat nějaký obecný "software".

Struktura příspěvku je následující. V části 2 uvedeme

základní ideje znalostního přístupu k výpočtům. V části 3 vysvětlíme myšlenku samo-zlepšujících se znalostních teorií a nastíníme mechanismy, které umožňují teoriím mít takové schopnosti. Hlavní výsledek, týkající se mezi znalostí inteligentních systémů, jsou formulovány v části 4.

2 Výpočet jako generování znalostí

V souladu se znalostní teorií výpočtů budeme nahlížet na znalosti jako na výsledek nějakého výpočetního procesu, který pracuje nad jistou znalostní doménou tak, že kombinuje její prvky — *elementární znalosti* — do 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.

V souladu se znalostní teorií výpočtů (viz např. [6], [7], [8], a [5]) budeme nahlížet na znalosti jako na výsledek nějakého výpočetního procesu. 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 či superinteligentní systémy jsou tudíž speciálními případy výpočetních systémů, u kterých je schopnost generovat znalost maximalizována. Ostatní procesy, které negenerují znalost, nebudeme považovat za výpočty. Zda narážíme na problém — jak poznáme, co je znalost? To je fundamentální filozofická otázka, na kterou filozofové doposud 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 *znalostní domény*, nad kterou výpočet operuje. Všechny znalosti o nějaké podmnožině znalostní domény jsou zachyceny pomocí *znalostní 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 strukturovanější, složitější 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 výpočtům je *strojově nezávislý*, protože platí pro jakýkoliv proces realizující odvozování v rámci dané znalostní teorie. Taktéž je *algoritmicky nezávislý*, protože se nezajímáme o to, jak je výpočetní pro-

ces 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 výpočtům umožňuje pracovat i takovými nedokonalé definovanými pojmy je předností našeho modelování. To nám umožní dosáhnout nové porozumění problematice generování znalostí které zatím nebylo možné dosáhnout jiným způsobem.

3 Inteligentní systémy se samo-zlepšujícími se teoriemi

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.

Takové vylepšení může v zásadě probíhat dvěma způsoby: *rozšířením*, kdy se k teorii přidávají nové znalosti, anebo *opravou*, kdy se odhalí "chyba" v teorii a následně se chyba odstraní. V prvním případě může být nová znalost přidaná či získaná "zvenč", od jiného inteligentního systému, anebo ji může systém získat sám, pomocí měření a experimentu. V druhém případě může systém odhalit chybu své teorie tak, že v teorii najde "znalostní kaz", tj. zjistí nesoulad teorie s pozorováním, nebo s jinou teorií nad stejnou znalostní doménou, anebo vygeneruje rozporné znalosti. Systém pak tuto chybu odstraní změnou vadné části teorie anebo dokonce změnou celé teorie. Systém, který má schopnost získávat nové znalosti právě popsánymi způsoby se nazývá *systém se samo-zlepšující se teorií*.

Podrobnější diskusi mechanismů, se kterými pracují samo-zlepšující se teorie, lze nalézt v plné verzi tohoto příspěvku [10].

Je zajímavé, že myšlenka nalezení "kazu" ve znalostní teorii je dobře známá ve filozofii počínaje dílem K. Poppera [4]. Popper modeloval pokrok ve vědě jako posloupnost po sobě následujících zamítání, *falzifikací* vědeckých teorií. Vpravdě,

naše myšlenka samo-zlepšujících se znalostních teorií modeluje Popperovu myšlenku v termínech znalostního přístupu k počítání a otevírá tak možnost pro její výpočetní realizaci.

Myšlenka samo-zlepšujících se znalostních teorií je v příkrém protikladu s myšlenkou tzv. samo-zlepšujícího (rozuměj: samo-zefektivňujícího) se software, která je obecně přijímána v teoretických kruzích umělé inteligence (viz např. [1], [11]). Podle názoru příslušných odborníků samo-zlepšující se software má být prostředkem, jak postupně zvyšovat inteligenci softwarových systémů tak, aby po čase překonali úroveň lidské inteligence. Naše modelování však zřetelně odhaluje, že klíčem ke zvyšování intelektuálního potenciálu inteligentních systémů není snaha o zvyšování jejich výpočetní efektivity (např. “vylepšováním” jejich inferenčních mechanismů), nýbrž soustavné zvyšování kvality a kvantity jejich znalostí. Klíčem jsou v první řadě znalostní data, nikoliv software, který s nimi pracuje.

4 Znalostní meze inteligentních systémů

Idea zvyšování inteligence výpočetních systémů pomocí samo-zlepšujících se teorií konkrétnějším způsobem potvrzuje kuhnovskou teorii [3] týkající se struktury vědeckých revolucí. V našem modelování odpovídá “vědecké revoluci” oprava vadné znalostní teorie. Mezi dvěma po sobě jdoucími revolucemi se věda vyvíjí tzv. *normálním způsobem*, tj. její vývoj probíhá pouze prostřednictvím zpřesňování již dříve dosažených znalostí. Tato období “revoluce” a normálního rozvoje se střídají, avšak je zřejmé, že intervaly zásadních změn, vyvolané zásadními objevy, se budou prodlužovat, až zásadní změny postupně vymizí.

Formálněji shrnuje toto poznání a jeho důsledky následující tvrzení o asymptotickém chování výpočetních procesů generujícím znalosti v dobře formalizovatelných znalostních teoriích:

Tvrzení: Ve stabilní dobře formalizovatelné znalostní doméně, jejíž podmnožiny popisují konečné znalostní teorie, v dlouhodobé perspektivě nastane situace, ve které

- všechny samo-zlepšující se znalostní teorie budou konvergovat k teoriím, generujícím ekvivalentní znalosti;
- dříve nebo později se každá taková teorie bude nadále vyvíjet jako tzv. *normální věda* (viz [3]), ve které již nastává rozšiřování ani opravování odpovídající znalostní teorie.

Jinými slovy, za uvedených předpokladů dosáhne poznání takového stupně, kdy již nelze očekávat žádné další převratné poznatky. Toto konstatování svým způsobem potvrzuje Horganovy ideje o “konci vědy”, které svého času způsobily značný rozruch ve vědeckém světě [2]. Nástin důkazu shora uvedeného tvrzení je ve ve finální verzi připravované práce [10].

5 Závěr

V příspěvku jsme ukázali nový pohled na výpočty, který je chápe jako procesy generující znalosti. Na rozdíl od klasického pohledu, který považuje za výpočty jakékoliv, i nesmyslné procesy, generované různými modely počítačů, náš přístup umožňuje zkoumat nejen kognitivní mechanismy produkce nových znalostních teorií, vedoucí k rozvoji inteligence

výpočetních systémů, ale také znalostní meze těchto generativních procesů.

Poděkování

Tento příspěvek vznikl za částečné podpory GA ČR v rámci grantové úlohy 15-049608, institucionálního plánu ÚI AV ČR RVO 67985807 a programu Strategie AV21.

Reference

- [1] Bostrom, N.: *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014
- [2] Horgan, J.: *The End of Science: Facing the Limits of Science in the Twilight of the Scientific Age*. New York: Broadway Books, 1996
- [3] Kuhn, T. S. *The Structure of Scientific Revolutions*. University of Chicago Press. 172 p., 1962
- [4] Popper, K.: *The Logic of Scientific Discovery*, Basic Books, New York, NY, 1959
- [5] van Leeuwen, J., Wiedermann, J.: Knowledge, representation and the dynamics of computation. To appear in: G. Dodig-Crnkovic, R. Giovagnoli (Eds): *Representation and Reality: Humans, Animals and Machines*, 2015, Berlin: Springer
- [6] Wiedermann, J. van Leeuwen, J.: Rethinking computation. In: *Proc. 6th AISB Symp. on Computing and Philosophy: The Scandal of Computation - What is Computation?*, AISB Convention 2013 (Exeter, UK), AISB, 2013, pp. 6-10
- [7] Wiedermann, J., van Leeuwen, J.: 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
- [8] Wiedermann, J., van Leeuwen, J.: What is Computation: An Epistemic Approach. (Invited talk). In Italiano, G.; Margaria-Steffen, T.; Pokorný, J.; Quisquater, J.J.; Wattenhofer, R. (ed.). *SOFSEM 2015: Theory and Practice of Computer Science*. LNCS 8939, Berlin : Springer, 2015, pp. 1-13
- [9] Wiedermann, J., van Leeuwen, J.: Towards a Computational Theory of Epistemic Creativity. In 41st Annual Convention of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB 2015). London: 2015, pp. 235-242
- [10] Wiedermann, J., van Leeuwen, J.: Getting to Grips with Superintelligence and AGI through the Epistemic Theory of Computations. Manuscript in preparation, April 2016
- [11] Yampolskiy, R.V.: On the Limits of Recursively Self-Improving AGI. In: J. Bieger (Ed.): *AGI 2015, LNAI 9205*, Springer, 2015, pp. 394-403

