IS 2019

Zbornik 22. mednarodne multikonference

INFORMACIJSKA DRUŽBA

Zvezek B

Proceedings of the 22nd International Multiconference

INFORMATION SOCIETY

Volume B

Kognitivna znanost

Cognitive Science

Uredili / Edited by
Toma Strle, Tine Kolenik, Olga Markič

http://is.ijs.si

10. oktober 2019 / 10 October 2019
Ljubljana, Slovenia

Zbornik 22. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2019

Zvezek B

Proceedings of the 22nd International Multiconference

INFORMATION SOCIETY – IS 2019

Volume B

Kognitivna znanost
Cognitive Science

Uredili / Edited by

Toma Strle, Tine Kolenik, Olga Markič

http://is.ijs.si

10. oktober 2019 / 10 October 2019
Ljubljana, Slovenia

Uredniki:

Toma Strle
Center za Kognitivno znanost,
Pedagoška fakulteta, Univerza v Ljubljani

Tine Kolenik
Odsek za inteligentne sisteme,
Institut »Jožef Stefan«, Ljubljana

Olga Markič
Filozofska fakulteta, Univerza v Ljubljani

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič
Na naslovnici je uporabljena slika robota podjetja 

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2019

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=302446848
ISBN 978-961-264-157-3 (epub)
ISBN 978-961-264-158-0 (pdf)

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2019

Multikonferenca Informaci družba (http://is.ijs.si) je z dvaindvajseto zaporedno prireditvijo tradicionalni osrednji srednjeevropski dogodek na področju informacijske družbe, računalništva in informatike. Informacijska družba, znanje in umetna inteligenca so - in to čedalje bolj – nosilci razvoja človeške civilizacije. Se bo neverjetna rast nadaljevala in nas ponesla v novo civilizacijsko obdobje? Bosta IKT in zlasti umetna inteligenca omogočila nadaljnji razcvet civilizacije ali pa bodo demografske, družbene, medčloveške in okoljske težave povzročile zadušitev rasti? Čedalje več pokazateljev kaže v oba ekstrema – da prehajamo v naslednje civilizacijsko obdobje, hkrati pa so notranji in zunanji konflikti sodobne družbe čedalje težje obvladljivi.

Letos smo v multikonferenco povezali 12 odličnih neodvisnih konferenc. Zajema okoli 300 predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic in 500 obiskovalcev. Prireditev bodo spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica (http://www.informatica.si/), ki se ponaša z 42-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2019 sestavljajo naslednje samostojne konference:

- 6. študentska računalniška konferenca
- Etika in stroka
- Interakcija človek računalnik v informacijski družbi
- Izkopavanje znanja in podatkovna skladišča
- Kognitivna znanost
- Kognitonika
- Ljudje in okolje
- Mednarodna konferenca o prenosu tehnologij
- Robotika
- Slovenska konferenca o umetni inteligenci
- Srednje-evropska konferenca o uporabnih in teoretičnih računalniških znanostih
- Vzgoja in izobraževanje v informacijski družbi

Soorganizatorji in podporniki konference so različne raziskovalne institucije in združenja, med njimi tudi ACM Slovenija, SLAIS, DKZ in druga slovenska nacionalna akademija, Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

V 2019 bomo sedmič podelili nagrado za življenjske dosežke v čast Donalda Michieja in Alana Turinga. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe je prejel prof. dr. Marjan Mernik. Priznanje za dosežek leta pripada sodelavcem Odseka za inteligentne sisteme Instituta »Jožef Stefan«. Podeljujemo tudi nagradi »informacijska limona« in »informacijska jagoda« za najbolj (ne)uspešne poteze v zvezi z informacijsko družbo. Limono je dobil sistem »E-zdravje«, jagodo pa mobilna aplikacija »Veš, kaj ješ?!«. Čestitke nagrajencem!

Mojca Ciglarič, predsednica programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

FOREWORD - INFORMATION SOCIETY 2019

The Information Society Multiconference (http://is.ijs.si) is the traditional Central European event in the field of information society, computer science and informatics for the twenty-second consecutive year. Information society, knowledge and artificial intelligence are - and increasingly so - the central pillars of human civilization. Will the incredible growth continue and take us into a new civilization period? Will ICT, and in particular artificial intelligence, allow civilization to flourish or will demographic, social, and environmental problems stifle growth? More and more indicators point to both extremes - that we are moving into the next civilization period, and at the same time the internal and external conflicts of modern society are becoming increasingly difficult to manage.

The Multiconference is running parallel sessions with 300 presentations of scientific papers at twelve conferences, many round tables, workshops and award ceremonies, and 500 attendees. Selected papers will be published in the Informatica journal with its 42-years tradition of excellent research publishing.

The Information Society 2019 Multiconference consists of the following conferences:

- 6. Student Computer Science Research Conference
- Professional Ethics
- Human – Computer Interaction in Information Society
- Data Mining and Data Warehouses
- Cognitive Science
- International Conference on Cognitonics
- People and Environment
- International Conference of Transfer of Technologies – ITTC
- Robotics
- Slovenian Conference on Artificial Intelligence
- Middle-European Conference on Applied Theoretical Computer Science
- Education in Information Society

The Multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM, SLAIS, DKZ and the second national engineering academy, the Slovenian Engineering Academy. In the name of the conference organizers, we thank all the societies and institutions, and particularly all the participants for their valuable contribution and their interest in this event, and the reviewers for their thorough reviews.

For the fifteenth year, the award for life-long outstanding contributions will be presented in memory of Donald Michie and Alan Turing. The Michie-Turing award was given to Prof. Marjan Mernik for his life-long outstanding contribution to the development and promotion of information society in our country. In addition, a recognition for current achievements was awarded to members of Department of Intelligent Systems of Jožef Stefan Institute. The information lemon goes to the “E-Health” system, and the information strawberry to the mobile application “Veš, kaj ješ?!” (Do you know what you eat?!). Congratulations!

Mojca Ciglarič, Programme Committee Chair
Matjaž Gams, Organizing Committee Chair

KONFERENČNI ODBORI
CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, Južna Afrika
Heiner Benking, Nemčija
Se Woo Cheon, Južna Koreja
Howie Firth, Škotska
Olga Fomichova, Rusija
Vladimir Fomichov, Rusija
Vesna Hljuz Dobric, Hrvaška
Alfred Inselberg, Izrael
Jay Liebowitz, ZDA
Huan Liu, Singapur
Henz Martin, Nemčija
Marcin Paprzycki, ZDA
Claude Sammut, Avstralija
Jiri Wiedermann, Češka
Xindong Wu, ZDA
Yiming Ye, ZDA
Ning Zhong, ZDA
Wray Buntine, Avstralija
Bezalel Gavish, ZDA
Gal A. Kaminka, Izrael
Mike Bain, Avstralija
Michela Milano, Italija
Derong Liu, Chicago, ZDA
Toby Walsh, Avstralija

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Marjetka Šprah
Mitja Lasič
Blaž Mahnič
Jani Bizjak
Tine Kolenik

Programme Committee

Mojca Ciglarič, chair
Bojan Orel, co-chair
Franc Solina
Viljan Mahnič
Cene Bavec
Tomaž Kalin
Jozsef Györkös
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Mitja Luštrek
Marko Grobelnik
Vladislav Rajkovič
Grega Repovš
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Tanja Urbančič
Boštjan Vilfan
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah

KAZALO / TABLE OF CONTENTS

Zbornik 22. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2019

Zvezek B

Proceedings of the 22nd International Multiconference

INFORMATION SOCIETY – IS 2019

Volume B

Kognitivna znanost
Cognitive Science

Uredili / Edited by

Toma Strle, Tine Kolenik, Olga Markič

http://is.ijs.si

10. oktober 2019 / 10 October 2019
Ljubljana, Slovenia

PREDGOVOR

Na letošnji konferenci Kognitivna znanost sodelujejo avtorice in avtorji z različnih disciplinarnih področij in predstavljajo tako empirične rezultate svojih raziskav kot tudi teoretska raziskovanja in razmisleke. Ena izmed osrednjih tem letošnje konference je "Umetna inteligenca in kognitivna znanost v 21. stoletju", avtorji pa se dotikajo tudi drugih področij kognitivne znanosti.

Upamo, da bo letošnja disciplinarno in metodološko bogata kognitivna konferenca odprla prostor za izmenjavo zanimivih misli in idej ter povezala znanstvenice in znanstvenike z različnih disciplinarnih področij, ki se ukvarjajo z vprašanji kognitivnih procesov.

Toma Strle
Tine Kolenik
Olga Markič

FOREWORD

At this year's Cognitive Science conference, the authors come from numerous disciplinary backgrounds and present their empirical as well as theoretical work. One of this year's main conference topics is "Artificial Intelligence and Cognitive Science in the 21st Century" but authors present research form other areas of cognitive science as well.

We hope that this year's cognitive conference – being extremely diverse in disciplines and methodologies – will become a welcoming space for exchanging intriguing ideas and thoughts as well as for bringing together scientists from all the different areas exploring the questions of cognitive processes.

Toma Strle
Tine Kolenik
Olga Markič

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Tine Kolenik,

Toma Strle

Olga Markič

Urban Kordeš

Matjaž Gams

Perception of linguistic and emotional prosody in Parkinson's disease - evidence from Slovene.

Maja Blesić
MEi:CogSci
Faculty of Education
University of Ljubljana
Slovenia
Majablesic2@gmail.com

Dejan Georgiev
Department of Neurology
University Medical Centre of Ljubljana
Slovenia
dejan.georgiev@kclj.si

Christina Manouilidou
Department of Comparative and General Linguistics
University of Ljubljana
Slovenia
Christina.Manouilidou@ff.uni-lj.si

ABSTRACT
The present study investigated the perception of emotional and linguistic prosodic functions in speakers of Slovene language affected by Parkinson's disease. Eight participants with a diagnosis of Idiopathic Parkinson's disease (PD group) and eight elderly healthy controls (HC), matched for age and years of education, were tested using an identification and a discrimination task for emotional and linguistic prosody. The stimuli for linguistic prosody consisted of sentences uttered as a question or as a statement. The stimuli for emotional prosody consisted of sentences uttered in six different emotional tones: anger, disgust, fear, happiness, sadness and pleasant surprise. Compared to healthy control the overall performance of the PD group was lower in three out of four tasks: linguistic identification, linguistic discrimination, and emotional discrimination. Moreover, the PD group identified less accurately negative emotions, more specifically anger and sadness.

Keywords
Parkinson's disease, receptive prosody, emotional prosody, linguistic prosody

1. INTRODUCTION
Prosody, the rhythm and melody of speech, plays many important functions in human communication. Through the variation of acoustic cues of pitch, loudness, and intensity speakers can convey linguistic (e.g. stress, sentence mode), as well as extra linguistic information (e.g. attitudes, emotions, irony and sarcasm) [1]. In neurolinguistics literature, two main functions of prosody are distinguished: linguistic and emotional [1]. Linguistic prosody encodes linguistic distinctions (e.g. phrase boundaries) [2]. Emotional prosody encodes information about the emotional state of the speaker or the emotional emphasis of the uttered content [3]. The processing of prosodic features of speech seems to rely on different neurocognitive mechanisms than the processing of other linguistic domains (e.g. syntax or semantics) [1]. A comprehensive model of the brain structures involved in emotional and linguistic prosodic processing is still missing [1]. Evidence from lesion [4] and neuroimaging [5] studies suggest that the basal ganglia, a subcortical structure with numerous connections to cortical areas, might play a role in how we process (express and perceive) linguistic and emotional prosody.

Parkinson's disease (PD) is a neurodegenerative disorder, characterized by the loss of dopaminergic cells in one of the nuclei of the basal ganglia. PD has been associated with expressive prosodic impairments and PD speech described as monotonous, lacking loudness and inappropriate in speech rate [6]. More recently, evidence for the receptive prosodic ability in PD has also been found [7-12]. Many studies investigating the perception of emotional prosody in PD reported a deficit in the recognition for specific emotions: sadness [11,12], anger [9], fear [9], and disgust [7,9,11]. Lower recognition rates in the perception of emotions in PD seem to converge on negative emotions [13]. Other studies [14-16] however, found no evidence for an impaired perception of emotional prosody in PD. Investigations of the recognition of linguistic prosody in PD report a preserved ability to recognize prosodic meanings of smaller units, such as words (e.g. PROject – noun, projECT – verb) and an impaired perception of prosodic meanings that require integration of prosodic information on longer units, such as for spoken sentences (e.g. the rising intonation indicating a question) [17]. The above described receptive prosodic difficulties in PD have been found independent of dementia or depression, but strongly correlated with executive functions and working memory capacity [8]. Among studies on prosodic disorders in patients with brain conditions, only few investigated the perception of both types of prosodies in the same group of patients. Moreover, contributions from Slavic languages are missing.

The aim of the current study was to investigate the perceptive ability of emotional and linguistic prosody on sentence level for speakers of Slovene language affected by PD, similarly to studies for Germanic (e.g. English; [10]) and Romance languages (e.g. Italian; [7]). For the investigation of linguistic prosody, we tested the identification and discrimination of questions and statements. For emotional prosody, we tested the identification and discrimination of utterances expressing six different emotional categories: anger, disgust, fear, happiness, sadness, and pleasant surprise. A prosody recognition paradigm consisting of a combination of an identification and a discrimination task was administered to the participants. Along the lines of Pell [10], we expected PD participants to perform less accurately in the linguistic and emotional identification tasks, but no impairment was expected in the discrimination task for linguistic and emotional prosody. Moreover, we expected the PD group to perform worse in the identification of negative emotions and the reduced recognition to be emotion specific.

2. MATERIALS AND METHODS

2.1 Participants

Eight individuals diagnosed with idiopathic PD (seven males and one female) and eight healthy controls (four males and four females), whose first language is Slovene, were included in the study. Participants of the PD group were recruited from the University Medical Center of Ljubljana, Department of Neurology. The participants for the HC group were recruited from the Retirement Home of Bežigrad, Ljubljana. Exclusion criteria for both groups included: dementia, hearing problems, language disorders, and depression. The neuropsychological assessment of participants included the administration of the Mini Mental State Examination (MMSE) [18]. The demographic data, together with the statistical comparison between groups using independent samples t-test, is presented in Table 1. The PD and HC groups did not differ significantly with respect to age $t(14) = -1.071$, $p = .370$ (PD = 77.38 ± 9.1; HC = 71.88 ± 11.3), years of education $t(14) = -1.007$, $p = .175$ (PD = 14.75 ± 4.6; HC = 12.75 ± 3.1), and MMSE scores $t(14) = 2.016$, $p = .063$ (PD = 28.13 ± 0.6; HC = 28.88 ± 0.8). Moreover, the comparison of the distribution of males and females between groups did not result as significant ($p = .282$, df = 1, Fischer's exact test).

Table 1: Demographic, neuropyschological, and neurological information for PD and HC (mean ± SD) together with the statistical comparison for age, years of education, and MMSE scores.

Variable	PD group Mean ± SD	HC group Mean ± SD	t-Test P value
Age (years)	77.38 ± 9.1	71.88 ± 11.3	> 0.05
Education (years)	14.75 ± 4.6	12.75 ± 3.1	> 0.05
MMSE (/30)	28.13 ± 0.6	28.88 ± 0.8	> 0.05

2.2 Materials

A new inventory of audio stimuli, uttered by an actress, was built for the purpose of this study. In order to ensure that the identification and discrimination would be based on prosodic cues and not on the content, pseudo-words (constructed from existing Slovenian syllables) were used in sentences (e.g. "*Prohast katoh groji zdrog")*. Ten raters first validated all stimuli. Included in the narrow selection were only those that scored high on the recognition test (at least 70%).

2.2.1 Stimuli-identification tasks

For the linguistic prosody identification task we used 20 utterances, 10 were statements and 10 questions. For the emotional prosody condition 42 utterances were used uttered in 6 distinct emotional tones: anger, sadness, disgust, fear, happiness and pleasant surprise (42 utterances: 7 utterances × 6 emotional categories).

2.2.2 Stimuli-discrimination tasks

The stimuli in the discrimination tasks consisted of pairs of prosodically same or different utterances. The content of two paired utterances was kept equal. For the linguistic prosody discrimination task 16 pairs of utterances were used, 8 uttered with the same and 8 with different intonation. For the emotional prosody discrimination task 20 pairs of utterances were used, 10 uttered with the same emotional tone and 10 with different.

2.3 Experimental tasks and procedure

For both experimental conditions (linguistic and emotional) we administered an off-line forced choice identification task followed by the corresponding off-line forced choice discrimination task. In the identification task single stimuli were presented in each trial (linguistic prosody condition: 20 trials; emotional prosody condition: 42 trials) and participants were asked to recognize and choose the correct label for stimuli belonging to distinct linguistic (question, statement) or emotional (anger, disgust, fear, happiness, sadness, pleasant surprise) categories. In the discrimination task, participants were presented with pairs of stimuli in each trial (linguistic prosody condition: 16 trials, emotional prosody condition: 20 trials) and were asked to judge whether they are the same or different in regard to prosody. To familiarize the participants with the tasks and speaker's voice, practice trials were presented before every task (not included in the analysis). Participants listened to the stimuli through headphones connected to a touch screen laptop on which they would give their responses.

2.4 Data analysis

Group differences between PD and HC in tasks were analyzed by comparing the proportions of correct responses (raw scores) to stimuli using the Chi-square test. Participant's responses (correct, incorrect) were in all comparisons treated as the dependent variable. The independent variables were: the two groups (PD and HC), the two different tasks (identification, discrimination), and the stimuli type in the identification tasks. The stimuli type for linguistic prosody were questions and statements. The stimuli type for emotional prosody were the six different emotional categories (anger, disgust, fear, happiness, sadness, and pleasant surprise), which were also grouped as positive (happiness and pleasant surprise) and negative emotions (anger, disgust, fear, and sadness).

3. RESULTS

Mean percentages of corrent answers of the PD and HC groups for linguistic and emotional identification and discrimination tasks are reported in Table 2.

Table 2: Mean percentages PD's and HC's correct responses in the identification and discrimination task for both conditions (linguistic and emotional prosody).

	Group	
Task	**HC**	**PD**
1. Identification		
Linguistic	94%	87%
Emotional	50%	43%
2. Discrimination		
Linguistic	93%	79%
Emotional	89%	80%

3.1 Linguistic prosody

Identification task: a significant difference between the participant's overall response to the stimuli was found $\chi 2(1, N = 320) = 5.297$, $p < .05$, with PD being less likely to respond

correctly (87%) compared to HC (94%) (see Table 2). No statistically significant differences in the response to questions $\chi2(1, N = 168) = 2.210$, $p = .137$ or statements $\chi2(1, N = 168) = 3.059$, $p = .080$ was found between groups. Discrimination task: a statistically significant difference $\chi2(1, N = 240) = 10.440$, $p < .01$ was observed between the groups in the overall proportion of correct responses, with PD performing worse (79%) compared to HC (93%) (see Table 2).

3.2 Emotional prosody

Identification task: no statistically significant difference between PD and HC was observed in their overall responses to the stimuli $\chi2(1, N = 672) = 3.449$, $p = .063$ (see Table 2). However, a comparison between PD's and HC's performance in response to negative emotions revealed a statistically significant difference $\chi2(1, N = 448) = 6.531$, $p < .05$, with PD (47%) scoring lower than HC (59%). No statistically significant difference was found between groups for positive emotions $\chi2(1, N = 224) = .183$, $p = .669$. Moreover, a comparison between PD's and HC's performance in response to specific emotions revealed a statistically significant difference for stimuli belonging to two emotional categories: anger $\chi2(1, N = 112) = 4.432$, $p < .05$ (PD 71%; HC; 87%), and sadness $\chi2(1, N = 112) = 10.351$, $p < .01$, (PD 37%; HC 68%). The mean percentage of PD and HC correct responses across different emotional categories is presented in Figure 1**.** Discrimination task: a statistically significant difference $\chi2(1, N = 320) = 4.073$, $p < .05$ was also observed in the overall correct responses between groups in the emotional prosody discrimination task, with PD performing worse (80%) compared to HC (89%) (see Table 2).

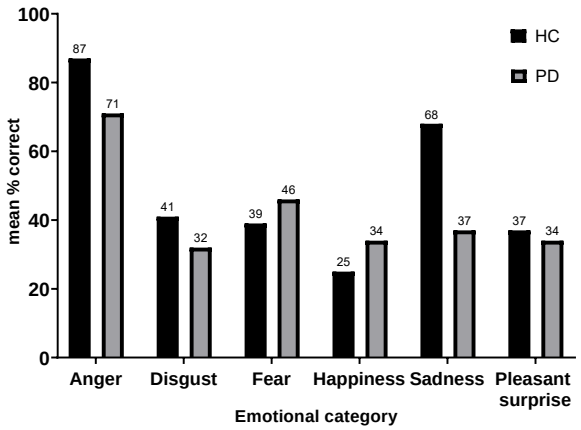


Figure 1: Mean percentage of PD and HC correct responses across the different emotional categories in the emotional identification task.

4. DISCUSSION

The present study sought to investigate the recognition of linguistic and emotional prosody in PD by providing evidence from Slovene language. Overall, compared to HC, the performance of the PD group was significantly lower in three out of four tasks: linguistic identification task, linguistic discrimination task, and emotional discrimination task. These findings did not confirm our first hypothesis, since we expected the PD group to perform significantly worse in the identification tasks only. Our findings are in contrast with Pell [10], where no low performance of PD in the emotional and linguistic discrimination task was found, but are in line with Ariatti, Benuzzi and Nichelli [7], who reported a low performance of PD in the discrimination tasks for both types of prosody. Moreover, Pell and Leonard [11] also reported a marginally significant worse performance of PD compared to HC in the discrimination of emotional prosody. Our PD group scored significantly lower than HC in the linguistic identification task, which tested the participants' ability to identify utterances as sentences or as questions based on intonation only, similarly to Ariatti et al. [7]. No statistically significant difference between PD and HC emerged in the overall scores in the emotional identification task. However, a further analysis comparing group performances in negative and positive emotions revealed a significant difference for negative emotions. The impoverished performance of PD was evident for the emotional categories of sadness and anger. These findings confirmed our predictions on PD's performance to be lower for negative emotions compared to positive ones and for it to be emotion specific. Our findings on low recognition rates for negative emotions (anger, disgust, fear, and sadness) and for the emotional categories of sadness and anger are in line with several other studies [9,11,12]. Overall, the results of our study supported the notion that PD affects receptive prosodic ability. Our study was the first attempt to investigate how Slovene speaking individuals diagnosed with PD perceive prosody conveying emotional and linguistic information on sentence level.

5. ACKNOWLEDGMENTS

We want to thank the actress Nataša Ulčar and the sound technician Brane Lenarčič for their immense help in developing the stimuli.

6. REFERENCES

[1] Kemmerer, D. L. 2015. *Cognitive neuroscience of language*, New York, NY: Psychology Press.

[2] Raithel, V. 2005. *The perception of intonation contours and focus by aphasic and healthy individuals*, Tübingen: Narr.

[3] Friederici, A. D., and Chomsky, N. 2017. Language in Our Brain. (January 2017). DOI= http://dx.doi.org/10.7551/mitpress/9780262036924.001.0001

[4] Cancelliere, A. E., and Kertesz, A. 1990. Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition* 13, 2 (1990), 133–147. DOI= http://dx.doi.org/10.1016/0278-2626(90)90046-q

[5] Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., and von Cramon, D. Y. 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping* 17, 2 (2002), 73–88. DOI= http://dx.doi.org/10.1002/hbm.10042

[6] Jones, H. N. 2009. Prosody in Parkinson's Disease. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders* 19, 3 (January 2009), 77. DOI= http://dx.doi.org/10.1044/nnsld19.3.77

[7] Ariatti, A., Benuzzi, F., and Nichelli, P. 2008. Recognition of emotions from visual and prosodic cues in Parkinson's disease. *Neurological Sciences* 29, 4 (2008), 219–227. DOI= http://dx.doi.org/10.1007/s10072-008-0971-9

[8] Breitenstein, C., Van Lancker, D., Daum, I., and Waters, C. H. 2001. Impaired Perception of Vocal Emotions in Parkinson's Disease: Influence of Speech Time Processing

and Executive Functioning. *Brain and Cognition* 45, 2 (2001), 277–314. DOI= http://dx.doi.org/10.1006/brcg.2000.1246

[9] Dara, C., Monetta, L., and Pell, M. D. 2008. Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions. *Brain Research* 1188 (2008), 100–111. DOI= http://dx.doi.org/10.1016/j.brainres.2007.10.034

[10] Pell, M. D. 1996. On the Receptive Prosodic Loss in Parkinson's Disease. Cortex 32, 4 (1996), 693–704. DOI= http://dx.doi.org/10.1016/s0010-9452(96)80039-6

[11] Pell, M. D., and Leonard, C. L. 2003. Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, and Behavioral Neuroscience* 3, 4 (2003), 275–288. DOI= http://dx.doi.org/10.3758/cabn.3.4.275

[12] Ventura, M. I., Baynes, K., Sigvardt, K. A., Unruh, A. M., Acklin, S. S., Kirsch, H. E., and Disbrow, E. A. Hemispheric asymmetries and prosodic emotion recognition deficits in Parkinson's disease. *Neuropsychologia* 50, 8 (2012), 1936–1945. DOI= http://dx.doi.org/10.1016/j.neuropsychologia.2012.04.018

[13] Péron, J., Dondaine, T., Le Jeune, F., Grandjean, D., and Vérin, M. 2011. Emotional processing in Parkinson's disease: A systematic review. *Movement Disorders* 27, 2 (September 2011), 186–199. DOI= http://dx.doi.org/10.1002/mds.24025

[14] Benke, T., Bösch, S., and Andree, B. 1998. A Study of Emotional Processing in Parkinson's Disease. *Brain and Cognition* 38, 1 (1998), 36–52. DOI= http://dx.doi.org/10.1006/brcg.1998.1013

[15] Martens, H., Van Nuffelen, G., Wouters, K., and De Bodt, M. 2016. Reception of Communicative Functions of Prosody in Hypokinetic Dysarthria due to Parkinson's Disease. *Journal of Parkinson's Disease* 6, 1 (2016), 219–229. DOI= http://dx.doi.org/10.3233/jpd-150678

[16] Mitchell, R. L., and Bouças, S. B. 2009. Decoding emotional prosody in Parkinson's disease and its potential neuropsychological basis. *Journal of Clinical and Experimental Neuropsychology* 31, 5 (2009), 553–564. DOI= http://dx.doi.org/10.1080/13803390802360534

[17] Pell, M. D., and Monetta, L. 2008. How Parkinson's Disease Affects Non-verbal Communication and Language Processing. *Language and Linguistics Compass* 2, 5 (2008), 739–759. DOI= http://dx.doi.org/10.1111/j.1749-818x.2008.00074.x

[18] Rakuša, M., Granda, G., Kogoj, A., Mlakar, J., and Vodušek, D. B. 2006. Mini-Mental State Examination: standardization and validation for the elderly Slovenian population. *Eur J Neurol* 2006;13:141–5. DOI = https://doi.org/10.1111/j.1468-1331.2006.01185.x

Mindfulness in preschool children – outline of the study

Tina Bregant
CIRIUS Kamnik
Novi trg 43 a
1241 Kamnik
++386 41 749 061
tina.bregant.drmed@gmail.com

Petra Plecity
Faculty of Education
Kardeljeva ploščad 16
1000 Ljubljana
++386 41 972 049
plecity_petra@hotmail.com

ABSTRACT

Practicing mindfulness with preschool and school-aged children affects their general well-being, diminishes mood swings, and improves their ability to focus which all contribute to effective learning. We present a research design for a study that is going to be carried out in the forthcoming months in selected kindergartens. The methodology is partially exploratory and partially following the methodology of the authors of Toy wrap Toy wait test.

We will try to establish whether preschool children become more focused on events around them and on their inner feelings while practicing mindfulness. We will test their ability to replace the dominant with subdominant reaction, which is one of the main components of self-regulation. On the basis of teachers' written answers, and questionnaires: a) Toy wrap Toy wait test and b) Children's Behavior Questionnaire, we will try to establish whether there is any difference between the children who practice mindfulness and those who do not, regarding their self-regulation.

Keywords

mindfulness, preschool children, kindergarten, Toy wrap Toy wait test, Children's Behavior Questionnaire

1. INTRODUCTION

Mindfulness has been known for thousands of years as a part of a meditative practice that, due to its specific way of focusing attention, allows it to focus on the present moment, thus calming the mind and reducing tension [1]. It can be used as a technique for psychotherapeutic purposes as it integrates content from cognitive, behavioral, experiential, and psychodynamic theories [2]. Practicing mindfulness in children affects their overall well-being and behavior, improves mood swings, helps with learning disabilities, fear of failure, and enhances executive function [3, 4].

1.1. Self-regulation

Self-regulation is a critical component of a child's readiness for school since it facilitates a child's acceptance by peers, social and academic success, higher self-confidence, professional achievements and better health [5]. Self-regulation is defined as the process by which people incorporate behavior change into their everyday lives, and it involves: self-monitoring, goal setting, reflective thinking, decision making, planning, plan enactment, self-evaluation and management of emotions arising as a result of behavior change. [6]. Self-regulation in childhood can be defined as a construct that represents the development of children's abilities to follow the everyday norms and practices that are embraced by their parents [7]. Self-regulation has been found to predict positive life outcomes, including good physical health (e.g., healthy body weight), higher levels of education and income, and better psychological well-being (e.g., lower risk for depression and substance abuse) [8]. Majority of the studies we had access to, focused on understanding children's maturation of executive functions —working memory (e.g. remembering a set of directions to complete a learning task), focused attention, and behavior inhibition (e.g. waiting for a turn to speak instead of talking out in class) and how these are linked to their development of emotional and/or behavior control during the preschool and early school years [9, 10].

To our knowledge, research on mindfulness in schools regarding the influence of mindfulness on self-regulation, is still in its infancy. Studies of mindfulness impact on behavior, academic performance, and physical health in children can best be described as 'promising' and 'worth trying'. There are data that show mindfulness training in pre-adolescence could support self-regulation development [11]. This is why we have decided to do the exploratory study where we are going to evaluate self-regulation in pre-school children who practice mindfulness compared to the control group.

1.2. Test Toy wrap Toy wait

Level of self-regulation represents the option to substitute your dominant response over subdominant one; to be able to avoid acting irrational and instead acting rational. A test named Toy wrap Toy wait [5, 11] establishes and compares the level of self-regulation in the intervention and control group.

The test named Toy wrap Toy wait is carried out in a way that the teacher tells the child that s/he has a surprise for him or her but first s/he has to wrap it. S/he sits the child down so that s/he is turned away from her/him by the angle of ninety degrees. The teacher starts to wrap the gift so that the crunching of the paper can be heard. After one minute the teacher shows the wrapped gift to the child and that is when the second part of the test begins. Researchers marks the latency of "peeping": seconds that pass before the child peeps and looks at the object the teacher is wrapping. When the teacher puts the wrapped gift in front of the child, s/he tells him/her to wait before s/he touches the gift and meanwhile s/he pretends that s/he has another task to do; s/he is tidying the paper from the previous task. Researchers marks the latency of touching the gift: how many seconds pass before the child touches the gift [12]. In this test the "peeping" and touching the gift represents the dominant versus subdominant response. In our case the dominant response is to "peep" straight away and the subdominant response is not to "peep" at all. The same reasoning is applied to touching the gift. Dominant response is to touch the gift straight away and the subdominant response is not touching it at all. Longer latency means better subdominant reaction and better self-regulation. The test does not need to be recorded. The teachers

are able to carry out the test and researchers can measure the latency of "peeping" and touching.

1.3. Children's Behavior Questionnaire

Children's Behavior Questionnaire is used for children aged from three to seven years. The questionnaire can be filled out by the parents. The questionnaire is a shorter version of the longer Children's Behavior Questionnaire [13] and its use was approved by the author. In the questionnaire three different subscales are used: liveliness (example: "is slow and unhurried in deciding what to do next"), negative emotion ("gets quite frustrated when prevented from doing something s/he wants to do") and effortful control ("notices it when parents are wearing new clothing"). Parents mark their child's behavior in the five-level Likert's scale, from extremely untrue to extremely true of their child.

2. METHODS

We will observe whether preschool children are capable of substituting their dominant reaction with the subdominant and whether they can react differently in their home environment by practicing mindfulness. In this study a dominant response will represent the response which is immediate, non-thinkable or irrational. On the other hand, subdominant response is a rational response, response that demands people to think first, before they act or respond. Our hypothesis is that with practicing mindfulness children will exhibit more subdominant responses than the control group measured by the test Toy wrap Toy wait and Children's Behavior Questionnaire.

The study is going to take place in two kindergartens in Municipality of Radovljica, Slovenia who have agreed to participate in the study. One group will contain 24 five to six-year-old preschool children in Kindergarten in Radovljica and the other one will contain same number, same age group preschool children in Kindergarten Lesce. Parents' approvals were collected prior to commencing the study.

The exercises of mindfulness practice will be carried out after lunch when children have time to rest. One department will practice mindfulness five times per week from five to fifteen minutes, eight weeks in a row while the control group will spend their time resting.

Teachers from individual kindergarten department will participate in the research. They will practice with children an eight-week mindfulness program. During the research they will be supported by the e-learning project of mindfulness: Shift Mindful, dare to be human. Mindfulness program for children will contain structured mindfulness practice based on different sources: mindfulness fairy tales, mindfulness activity games, mindfulness tasks, such as focus on their breathing, focus on gratitude or different feelings; experience love, anger, sadness, happiness etc. Teachers in the control group will follow the regular curriculum and will read a story to children when they rest or let them play quietly.

The research will be composed of three parts. The first part of the research will present a test named Toy wrap Toy wait [5] which will establish and compare the level of self-regulation in the intervention and control group. The second part of the research will present the shorter version of Children's Behavior Questionnaire [13] for children aged from three to seven years which will be filled out by the parents. Parents will get the Questionnaire in hand and they will fill it out before and at the end of the research. With the Questionnaire we will get the report on children's behavior at home, with special focus on concerning subscales; liveliness, negative emotion and effortful control. Our intention is to verify possible changes happening in behavior of children in their home environment while practicing mindfulness. The third part of the research will present the teachers' answers who will carry out the exercises of mindfulness. We will ask them three questions concerning the mindfulness practice with children. We will be interested on their opinion on mindfulness in general, what do they think about practicing mindfulness with preschool children and whether they noticed any change in children. The questionnaire will contain written questions and answers send via email.

Table 1: Tests used in the experiment

Test	Time	Short description
Toy wrap Toy wait	5 minutes	Wrapping the gift by the teacher in front of the child and measuring the time when peeping and touching from the child starts
Children's Behavior Questionnaire	15 minutes	Parents have to mark their child's behavior in the five-level Likert's scale, from extremely untrue to extremely true of their child
Questionnaire for the teacher	10 minutes	Written answers about experience in practicing mindfulness and its effect on children

3. DISCUSSION

3.1. Ethical concerns

The practice of mindfulness and meditation is a conscious exercise that builds attention control and inhibitory skills [10]. Recent pilot research on practicing mindfulness has shown a positive effect on the general well-being, behavior, improving mood swings, and help with learning problems, fear of failure and strengthening executive functions [14-16]. Teaching children mindfulness is expected to enhance their regulatory competences and give them a new experience. The technique is itself non-invasive, voluntary, and can be seen as part of child play. The tests used are playful so children are not stressed by doing them. Since we already introduced the research to the teachers in the kindergarten where the research will take place, they already expressed their interest since they lack the techniques which could help them concentrate and focus themselves.

3.2. Limitations

One has to be aware that incorporating more integrative therapies or techniques in preschool and school programs could be motivated by some economic / consumer interest. However, mindfulness is a technique which requires little financial input. Another problem is that the structure of teaching mindfulness can vary substantially from teacher to teacher, which is why we have decided to offer the teachers a uniform course on mindfulness. Another major drawback to our study is the lack of time. Practicing mindfulness techniques takes more time to exert any larger and/or measurable effects. Practicing mindfulness is most valuable when an individual can internalize or in-personalize the practice in her or his daily routine which can be done by practicing mindfulness on daily basis, for a longer time period. The sample in our study could be too small, since it will be composed of two groups of preschool children, whose parents will agree to participate in the two selected

kindergartens in the Municipality of Radovljica. In our study we will not include children's personality and socio-economical and emotional (family) background, which are all important components in accepting mindfulness practice in daily routine.

4. CONCLUSIONS

To introduce practicing mindfulness into kindergartens and primary schools as a part of the curriculum or as a part of a learning program could bring some positive effects on children which are crucial for each individual who starts the path of public and private educational establishment. One of the positive effects could be a higher level of self-regulation. Hopefully with our study we will be able to show some effect on preschool children who will practice mindfulness with the help of their kindergarten teachers.

5. ACKNOWLEDGMENTS

Our thanks to Vesna Laković, the cofounder of the project in Serbia, entitled Shift Mindful, dare to be human, a mindfulness program, which can be carried out by mentoring via IT use.

6. REFERENCES

[1] Bregant, T. 2015. Čuječnost za odrasle, ki delajo z otroki: kako polno zaživeti. *Didakta.* 25, 180-182, 54-57.

[2] Brown, K.W. and Ryan, R.M. 2004. Perils and promise in defining and measuring mindfulness: observations from experience. *Clin. Psychol. Sci. Pract.*, 11, 242 – 248.

[3] Flook, L. 2010. Effects of Mindful Awareness Practices on Executive Functions in Elementary School Children. *Journal of Applied School Psychology.* 26,70–95. DOI=http://dx.doi.org/10.1080/15377900903379125

[4] Flook, L. et al. 2015. Promoting Prosocial Behavior and Self-Regulatory Skills in Preschool Children Through a Mindfulness-Based Kindness Curriculum. *Developmental Psychology*, 51, 1, 44- 51.

[5] Moffitt, T.E. et al. 2011. A gradient of childhood self-control predicts health, wealth and public safety. *Proceedings of the National Academy of Sciences*, 108, 7, 2693- 2698.

[6] Laranjo, L. 2016. Social Media and Health Behavior Change. *Participatory Health Through Social Media*, 83-111. DOI= https://doi.org/10.1016/B978-0-12-809269-9.00006-2

[7] Kopp, C. B. 2001. *International Encyclopedia of the Social & Behavioral Sciences*, pp. 13862-13866. DOI= https://doi.org/10.1016/B0-08-043076-7/01775-7

[8] Diamond, A. 2016. Why improving and assessing executive functions early in life is critical. In: *Executive Function in Preschool Age Children: Integrating Measurement, Neurodevelopment and Translational Research.* Washington, DC: American Psychological Association.

[9] Diamond, A., and Lee, K. 2011. Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333(6045), 959-964.

[10] Kaunhoven, R.J., and Dorjee, D. 2017. How does mindfulness modulate self-regulation in pre-adolescent children? An integrative neurocognitive review. *Neurosci Biobehav. Rev.*, 74, 163-184. DOI=https://doi.org/10.1016/j.neubiorev.2017.01.007

[11] Razza, R.A., Bergen- Cico, D., Raymond, K. 2013. Enhancing Preschoolers' Self. Regulation Via Mindful Yoga. Springer. *J Child Fam Stud*. DOI=http://dx.doi.org/10.1007/s10826-013-9847-6

[12] Murray, K. T., and Kochanska, G. 2002. Effortful control: Factor structure and relation to externalizing and internalizing behaviours. *Journal of Abnormal Child Psychology*, 30, 5, 503–514.

[13] Putnam, S. P., and Rothbart, M. K. 2006. Development of short and very short forms of the children's behavior Questionnaire. *Journal of Personality Assessment*, 87, 1, 102–112.

[14] Cillesen, L. 2016. *Mindfulness in Adolescents with Asthma: Role in Quality of Life and Asthma Control in an Observational and a Treatment Study.* Master Thesis. Radboud University Nijmegen.

[15] Lo, H. et al. 2016. The effect of a family-based mindfulness intervention on children with attention deficit and hyperactivity symptoms and their parents: design and rationale for a randomized, controlled clinical trial (Study protocol). *BMC Psychiatry.* 16:65. DOI=http://dx.doi.org/10.1186/s12888-016-0773-1

[16] McClafferty, H. 2018. Mind-Body Therapies in Paediatrics. *Alternative and Complementary Therapies*, 24, 1. DOI=http://dx.doi.org/10.1089/act.2017.29143

Consequences of relationships with robots in our everyday lives

Izabela But
Faculty for Education
Kardeljeva ploščad 16
1000 Ljubljana
+386 51 664369
Izabelabut92@gmail.com

ABSTRACT
We live in an era where robotics and artificial intelligence are rapidly developing, resulting in first humanoid robots. A novelty like this could have a great impact on the society. It is therefore important to consider what potential positive and negative consequences the introduction of humanoid robots into society might pose. At first, this article briefly presents a pilot study on attitudes of the elderly towards robots in our everyday lives. We will also consider a similar research. We used both to present general attitudes towards robots. We will then continue with further investigation on how and why the robots influence humans – we will point out some of the human and the robot characteristics that are involved in their relationships and discuss why some consequences, regarding characteristics, are good or bad for human beings.

Keywords
artificial intelligence, robotics, robots, humanoid robots, ethical issues

1. INTRODUCTION
Every novelty in science, that will soon penetrate/infiltrate our everyday lives, should be carefully considered, because we must be prepared for the consequences – the good and the bad, and how they could possibly influence the society and individuals. In this article we will talk about robots that are entering our day-to-day lives in many different ways; from the industry, and the army to our homes. We will start with my pilot study and a similar research. They will serve as a starting point for further discussion, where we will identify relevant characteristics of humans and robots, and how those influence their relationship.

2. PILOT STUDY
The main aim of the study was to investigate what the elderly think about the usage of robots and how robots make them feel. Hypothesis is that the elderly are not as open, to novelties, such as robots in our everyday lives, as young people are. Coincidentally, participants in both groups differed in level of education and because of that there is another hypothesis; highly educated participants will be more open minded for novelties and will consider them more critically than the ones with lower education.

2.1 Participants
There were two groups of participants. Both consisting of four women. In the first group, the participants were 80 years old or more, and in the second group the age span was from 70 to 80 years old.

Because of a fortunate coincidence, participants in the first group were highly educated – they have all completed university-level education. Participants in second group have not. This information was used to form the second hypothesis, presented a few lines above.

2.2 Method
Before we started the interview, I asked all the participants about their usage of computers and cellphones. This information showed whether they were familiar with some forms of technology.

All participants in the first group had their own phones, but not a computer although they did use computers in the later years of their careers. They admitted that it made some work easier, but not all of it. For example, it was really helpful when you needed a calculator, but not useful in making compromises with clients or selling the products. Today they are using computers for writing e-mails and looking up information on the world-wide net. Only one of them is avoiding computers and prefers newspaper and books. On the bases of their education and the use of computers and phones I think they will be open but critical to new technology.

In the second group, there were four elderly ladies from a smaller town. They didn't have such high education as the participants in the first group. They were a cook, two maids in the local hotel and a cashier at the local market. They have lived in their home town for all their lives and they never travelled. During their careers, they were not in contact with computers. Today, they have their own cellphones, but not any computers. However, they all have TVs, for which they said are a good source of information for them. They said they don't have any need to learn how to use a computer or a smart phone. As we can see, the second group is quite different from the first one. My hypothesis is that they won't be as open for new technology, as the participants in the first group.

After this introduction we proceeded with the interview, based on some videos from the portal "Youtube" [1,2]. Questions were prepared beforehand. Questions were as following: "How do you feel about robots helping you in your work place? Would you trust them with your duties? Do you think we should welcome industry robots in our work places if the human workers won't lose their jobs because of it?", "Pepper is a social robot; you can talk to her, play cognitive games with her, she can make you feel less lonely. Would you accept this kind of robot in your home?", "If you'd live in a home for the elderly would you prefer a robot or a human helper and caregiver?", "We can use robots for human rehabilitation. Would you be open to trying it or would you prefer

a human physiotherapist?", "Do all these robots we have seen in videos make you feel good and safe, or are you having any doubts and why so? What do you like about these robots and what makes you feel uncomfortable? Do you think they would be more successful in social interactions because they would constantly be in a good mood and would be made to satisfy humans?"

2.3 Results

In the first part of the first video [1] we can see an industrial robot. General manager said that they thought the robots were good for their company because they didn't need rest and they represent a lower expense than a human yearly salary. Besides, human employees had accepted them, as well. In the first group, participants agreed that it made sense to employ robots in the industry, if it will be taken care of the human workers left without jobs. Through discussion we came to an idea of a universal salary for all people, but came to a problem that owners of such companies would not give up on "easy money" easily. And as consequence, the rich would become even richer and the poor even poorer. To a question, if they would have a robot helper at their work place, they answered differently. But what was common to all of the answers was that they would miss the human factor. Hence, they would have a robot for things like math, but not for something where human factor is important. They also expressed doubt about robots' lack of plasticity and ability to adapt to situations. Participants in the second group agreed that robotization would relieve human employees. Robots are also faster and more precise than humans. In the second group, we also came to a problem of universal salary. To the second question, they all answered negatively and argued the same way as the participants in the first group – robots lack human factor.

Second part of the first video [1] was showing a robot named Pepper and her interaction with her owners. I wanted to know if they would have such a robot in their home. All participants in the first group expressed doubt about having a robot as a friend or a companion. It would feel odd to have a robot friend, again because of their lack of "human-ness". The answers in the second group were not as I had expected. Three of the participants were widows and they all answered that they would like to have such a companion, but at the same time, they expressed doubt about how they would use it, because they didn't know anything about robots.

Third part of the video [1] shows a robot which is used as outer skeleton for patients in the process of rehabilitation after a stroke or some other accident. Robot is connected to the patient through electrodes on muscles and it helps the patient walk. After therapy, the odds of walking again get higher. Participants in both groups said that it was a really useful robot. Regarding this robot we also talked about robot-surgeons; would they prefer a human or a robot surgeon? They said that a robot could be more precise, but they were doubtful about its ability to decide fast if something would go wrong.

In the next video [2] we saw a robot similar to Pepper. It was working as a companion and an animator for exercise in a home for the elderly. Participants in the first group weren't that eager about a robot employee in a home for the elderly. In the second group, they said it would be fun to have additional staff that would be robots.

The last set of questions was about general impressions of robots in our everyday lives. Participants in the first group were sceptic about the universality of robots and missing the human touch, which they found very important in human communication. Participants the in second group came to a similar conclusion. Though, as they said, they didn't have enough knowledge and understanding to judge this.

2.4 Discussion

This pilot study had two groups with four participants each. This means the sample is very small and unrepresentative. This is why the results cannot be used for the whole population of the elderly. Though this study was useful for the purpose of making a starting point for further investigation.

Participants in the first group all had a university education. while the participants in the second group did not. My hypothesis was that the level of education will influence how participants comprehend the robots and its use. Study showed what I had expected; participants in the first group were more open to the usage of the robots. What people think of usage of robots also depends on the culture and where they live. Robots are well accepted in Japan, but in Europe there are still some doubts about it. What could also influence the results is the fact that all participants live alone in their own homes, and not in homes for the elderly. I think people in homes for the elderly often feel neglected by their family and have a bigger need to have someone close to them, even if it is a robot. Because I didn't have access to database, my sample was not random, but I used participants I or my mentor knew.

The aim of this pilot study was not to generalize the results but more to get a grip about the attitudes of the elderly towards robots and their presence in our everyday lives.

3. SIMILAR RESEARCH

Dautenhahn with colleagues [3, p. 1] made a similar research as the one presented before. They were investigating attitude towards potential robot companions. Their main aim was to figure out, how people perceive robots and how they feel about their presence in our everyday lives.

In their research there were 28 participants. Their research questions were: "Are people accepting of the idea of robot companions in the home?", "What are people's perceptions of a future robot companion?", "What specific tasks do people want a robot companion to perform?", "What appearance should a robot companion have?", " What are peoples' attitudes towards a socially interactive robot in terms of robot behaviour and character traits?", "What aspects of social robot-interaction do people find the most and least acceptable?". [3, p. 2].

They used two types of questionnaires: the Cogniron Introductory Questionnaire, used for providing demographic details and the Cogniron Final Questionnaire used for investigating people's attitudes and perceptions towards robots. First questionnaire enquired about participants' personal details (age, gender, occupation), level of familiarity with robots, prior experience with robots (at work, as toys, in movies/books, in TV shows, in museums or in schools), and level of technical knowledge of robots were rated according to a 5-point Likert scale. And the second consisted of questions like "What is robot companion?", "What tasks would you like a future robot to be able to carry out?", "How controllable, predictive and considerate should a future robot be?", "How human-like should the robot appear, behave and communicate?", etc. [3, p. 2-3].

Results showed that 82% of subjects liked or liked very much the concept of computing technology in the home compared to just under 40% for a robot companion. When asked what role they thought a future 'robot companion in the home should have', the majority of participants wanted the robot as an assistant (79%), a machine/appliance (71%) followed by a servant (46%). Younger participants even said they would have robot as a friend and companion. Majority would like future robots to carry out household job as vacuuming. Only 10% would trust a robot with babysitting. Most participants expressed that they would want the behaviour of a robot companion to be highly predictable. Participants' responses about human-like appearance, behaviour and mode of communication for a robot companion were somewhat mixed. 71% of subjects would want a robot companion to communicate in a very human-like or human-like manner. However, human-like behaviour and appearance were less desirable. 36% thought that the robot should behave either very human-like or human like, and 29% stated that a robot in the home should appear human-like or very-human like. [3, p. 3-4].

Suma sumarum; Most subjects saw the potential role of a robot companion in the home as being an assistant, machine or servant. Few were open to the idea of having a robot as a friend. Robot companions should also be predictable, controllable, considerate and polite. Their communication should be human-like, though their appearance and behavior are not necessarily human-like. [3, p. 4]

The current study was exploratory in nature and has revealed many findings that could be relevant for future research ideas and robot companion designs. However, a potential drawback of the study could be the self-selected university sample that was recruited to participate. Future studies should attempt to recruit a more representative population sample. Also, the cultural background of subjects, which was not accessed in the present study, is likely to have a significant impact on people's perception of robots. Moreover, none of the participants were older than 55 years, which means that
the views of an elderly population are likely to be under represented in this study. [3, p. 4]

To conclude, the current study explored people's perceptions and attitudes towards the idea of a robot companion in the home. Interesting and positive results have emerged, indicating that a large proportion of people are favourable to the idea of a robot companion. Results have highlighted the specific roles and tasks that people would prefer a robot companion to perform in addition to the desired behavioural and appearance characteristics. The finding that people frequently cited that they would like a future robot to perform the role of a servant is maybe similar to the human 'butler' role [3, p. 5-6].

4. COMPARISON OF BOTH RESEARCH

Both researches had a similar goal: create a picture of human attitude towards robots and their presence in our everyday lives.

In both researches participants found robots to be acceptable for carrying out household jobs. At the same time, they all rejected the idea of robot as a friend.

Regarding both researches I think people don't accept robots as substitutes for human beings, although they are already taking our jobs in the industry, help in our homes, hospitals, hotels, etc.

5. HUMANS AND ROBOTS

Both researches gave us an insight on attitude of humans toward robots. Now we can continue discussing about our relationship with robots as our partners, friends or lovers and how could this relationship affect humans and society. Relationship depends on characteristics of both groups.

Human characteristics that influence our relationship with robots are: emotions and the ability to anthropomorphize, which is the ability to see non-living things as living. Itis a psychological characteristic that we got through evolution. In this process human ascribe human characteristic to non-human objects or subjects. Emotions are, like the ability to anthropomorphize, a part of human cognition. We got them during evolution and they are helping us regulate our living in day-to-day lives. The consequence of both is that a human being bonds emotionally with a robot very quickly. This could be ethically problematic because such a relationship is only one-sided. This is why friendships or partnerships with robots could be ethically problematic.

There are also robot characteristics that influence the relationship: mobility, autonomy, way of communication. I will present a few experiments on how autonomy and mobility of the robot influence human perception of them. Regarding human psychology and robots' construction and mechanics we can get to another ethical problem: loss of tolerance towards another human being.

5.1 Autonomy and mobility experiments

Scheutz [5, p. 208] was doing a research on how autonomy and mobility influence human perception of robots. Autonomy is considered as the ability to carry out a task without human intervention. And there can be different levels of autonomy. We can give orders to a robot such as "Move 3m ahead" or "Find the evidence for stratification in this rock". It is obvious that the robot that can carry out the second order, has a higher level of autonomy. Levels differ among them, depending on the ability of comprehension, analytics, communication, decision making … [5, p. 208]. Scheutz made three different experiments.

5.1.1 Dynamic Autonomy

In this task, a human subject worked together with a robot to accomplish a team goal within a given time limit. While both the human and the robot had tasks to perform, neither robot nor human could accomplish the team goal alone. In one of the task conditions (the "autonomy condition"), the robot was allowed to act autonomously when time was running out in an effort to complete the team goal. As part of this effort, it was able to refuse human commands that would have interfered with its plans. In the other condition (the "no autonomy condition"), the robot would never show any initiative on its own and would only carry out human commands. Human subjects were tested in both conditions (without knowing anything about the conditions) and then asked to rate various properties of the robot. Overall, subjects rated the "autonomous robot" as more helpful and capable, and believed that it made its own decisions and acted like a team member. There was also evidence that they found the autonomous robot to be more cooperative, easier to interact with, and less

annoying than the nonautonomous robot. Surprisingly, there was no difference in the subjects' assessment of the degree to which the robot disobeyed commands (even though it clearly disobeyed commands in almost all subject runs in the autonomy condition while it never disobeyed any commands in the no-autonomy condition). We concluded that subjects preferred the autonomous robot as a team partner. [5, p. 209]

The problematic point of this relationship between human and a robot is, that it is one-sided. Robots are not capable of forming emotional bonds or feeling emotions. At this moment, they are only capable of recognizing human emotion and act accordingly-depends on how they are programmed. In my opinion, genuine features of partnerships or friendships are reciprocity of emotions and respect and belonging. This makes a human happy and fulfilled. Today, robots are not as sophisticated and developed to be able to feel the emotions or be capable of forming an emotional bond with its owner. Because of that, the relationship with a robot cannot be as good as the relationship with a living being. If humans, instead of a robot, buy a dog, this relationship will fulfill reciprocity.

5.1.2 Affect Facilitation

Here, instead of making autonomous decisions, the robot always carried out human orders. However, in one condition (the "affect condition") it was allowed to express urgency in its voice or respond to sensed human stress with stress of its own (again expressed in its voice), compared to the "no-affect condition," where the robot's voice was never modulated. Each subject was exposed to only one condition and comparison was made among subject groups. The results showed that allowing the robot to express affect and respond to human affect with affect expressions of its own—in circumstances where humans would likely do the same and where affective modulations of the voice thus make intuitive sense to humans—can significantly improve team performance, based on objective performance measures. Moreover, subjects in the "affect condition" changed their views regarding robot autonomy and robot emotions from their pre-experimental position based on their experience with the robot in the experiment. While they were neutral before the experiment as to whether robots should be allowed to act autonomously and whether robots should have emotions of their own, they were slightly in favor of both capabilities after the experiments. This is different from subjects in the no-affect group who did not change their positions as a result of the experiment. We concluded that appropriate affect expression by the robot in a joint human–robot task can lead to a better acceptability of robot autonomy and other human-like features, like emotions in robots. [5, p. 209-210]

5.1.3 Social Inhibition and Facilitation

While the previous two studies attempted to determine human perceptions and agreement with robot autonomy indirectly through human participation in a human–robot team task (where the types of interactions with the robot were critical for achieving the goal, and thus for the subjects' views of the robot's capabilities), the third study attempted to determine the human-likeness of the robot directly. Specifically, the study investigated people's perceptions of social presence in robots during a sequence of different interactions, where the robot functioned as a survey taker as well as an observer of human task performance. Our experimental results showed that robots can have effects on humans and human performance that are otherwise only observed with humans. Interestingly, there was a gender difference in subjects' perception of the robot, with only males showing "social inhibition effects" caused by the presence of the robot while they were performing a math task. Post-experimental surveys confirmed that male subjects viewed the robot as more human-like than did the female subjects. [5, p. 210-211]

The results showed human attitude toward autonomous robots. People prefer autonomous robots, when they have to finish the task together. Humans prefer characteristics that shows human-like autonomy. It is important to acknowledge, that this might not be the case in a situation outside the laboratory. Let us now check the situation outside the laboratory.

5.1.4 Robots, mines and soldiers

Now we will talk about a robot that is used for detonating the mines. It goes over the dangerous mine field and when it steps on it, the mine blows up/explodes. The robot was made by Mark Tilden who was present in an experiment. Every time the robot found a mine, it was left with less and less limbs. When it only had one left, it was still pulling itself forward. Then, Tilden stopped the experiment saying he could not stand the pathos of watching the burned and crippled machine drag itself forward. This test is in his opinion inhumane. [5, p. 211]

Whether or not "inhumane" was an appropriate attribution, the fact remains that the only explanation for not wanting to watch a mindless, lifeless machine, purposefully developed for blowing up mines, destroy itself, is that the human projected some agency onto the robot, ascribing to it some inner life, and possibly even feelings. [5, p. 211]

We can conclude that the more sophisticated the robots get, the bigger will be the danger for humans to form one-way emotional bond with such robots. One-way emotional bonds are potentially dangerous because we could be doing things we otherwise wouldn't. For example: if we would trust robots too much, it could get us to buy some articles we don't need, just because it said so. And it could say so, if it were programmed this way. [5, p. 216] I also think one-way emotional bonds are harmful for people who bond this way. Relationships we have should be reciprocal, because this gives the fullness and depth to the relationship.

People who are selling robots should inform their clients that robots don't have emotions and cannot form emotional bonds. This way, they can instill knowledge about non-reciprocal relationships.

Robots are made to make our lives easier and better, which doesn't mean there cannot be some bad consequences. This is why it is very important to think about all possible outcomes of having a robot in our home. And because of the possible negative consequences we should also prepare some safeguards. These safeguards could be laws or guides on how robots can be made, and obligatory informing of clients that robots don't have emotions and cannot bond this way. Which still doesn't prevent us from bonding to robots. [5, p. 217-218]

To conclude, Scheutz approached the problem with doubt in such robots and with a lot of criticism. I think his way of thinking makes sense because society is not informed enough and not everyone is educated on the topic, or they don't even think about negative consequences. Usually people and society are so fascinated by the achievements of science, they forget to think about the bad consequences. We should somehow prevent that.

5.2 Partnership with robots

At some point, the robots could also become partners and lovers. I think partnership is one of the most important relationships we have in our lives. Partner (husband/wife) is someone you supposedly spend the rest of your life with. We choose our partners in many different ways, by different criteria: regarding looks, personality traits, goals, way of communication … What is also important in partnership is reciprocity of respect and emotions. Because of what we have said until now, we can easily claim that a robot would not make a good partner. Downside of having a robot for a partner is also that they are not equal to us; we chose them, they are made the way we want them to be, we don't have to compromise with them, because they always agree with us, etc. Because of how robots function and how they influence humans and our perception of human beings and relationships, partnership with robots could bring us more bad consequences than good. I think the only good outcome would be that the person wouldn't be alone. Otherwise it would change our perception on how relationships work: it is possible that humans would lose patience towards another human being, their potential partner, because they would be used to not compromising. Also, other humans don't think the same way as we do, and have different goals and taste in different things in our lives. Robots would support the fact that we don't have to work for a relationship.

We could make criteria on who is justified to have a robot as a partner, but how would it look? Will the justified be someone who got dumped by his or her first girlfriend/boyfriend? Someone who got divorced for the second time? Or someone who is working 12 hours a day and doesn't have time for social interaction? There are many different questions which could help us define these criteria, but how will we choose the best one? This could be a topic for a whole another article, so I will end it here.

I argue that having a robot partner or a lover is not good.

First, it can clearly be argued that a peaceful, even loving interaction among humans is a moral good in itself. Second, we should probably distrust the motives of those who wish to introduce technology in a way that tends to substitute for interaction between humans. Third, for a social mammal such as a human, companionship and social interaction are of crucial psychological importance. Ultimately, it may perhaps be that we can scientifically analyze all of these psychological needs. It may also be possible one day to build technology that completely fulfills these needs. However, as things stand, we cannot be sure that our caring technologies are capable of meeting all the relevant psychological needs. [3, p. 238]

6. CONCLUSION

Robots are a part of our cultural and technological evolution. It is only a matter of time before they will infiltrate our society completely. I think the right time to prepare ourselves for that moment is now. I think all the scientists that are included in producing a robot should think about how such robots will influence the society. I also think the philosophers should help them think and rethink all the possible outcomes and consequences and how we could prepare for them or even prevent them.

Humans and robots are two different categories, and each have different characteristics which influence one another. We have to consider all of them, when we think about how the relationship among them will work.

In this article, I first presented my pilot study. The main aim of the study was to get a grip on how the elderly feel about robots in our everyday lives. Results confirmed my first hypothesis. Regarding second hypothesis, I was wrong in suggesting that better educated participants would be more open to having a robot in their home. After presenting my pilot study, I also presented a few other studies considering human relationship with robots.

I finished this article with the thought of why robots are not good for us as partners.

7. REFERENCES

[1] SBS Dateline. 2017. The Japanese robots used for companionship, household tasks and sex. *Youtube*. Retrieved September, 2019 from https://www.youtube.com/watch?v=YzzDLujpat4&t=207s

[2] Nine News Perth. 2017. Robot Therapy | 9 News Perth. *Youtube*. Retrieved September, 2019 from https://www.youtube.com/watch?v=FZegHxMimMk

[3] Whitby, B. 2012. *Do You Want a Robot Lover? The Ethics of Caring Technologies.* Robot Ethics: The Ethical and Social Implications of Robotics. MIT, Massachusetts, 233-246.

http://kryten.mm.rpi.edu/Divine-Command_Roboethics_Bringsjord_Taylor.pdf

[4] Dautenhahn, K. 2005. *What is a Robot Companion - Friend, Assistant or Butler?* 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems.

http://homepages.herts.ac.uk/~comrklk/pub/Dautenhahn.etal_IROS05.pdf

[5] Scheutz, M. 2012. *The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots*. Robot Ethics: The Ethical and Social Implications of Robotics. MIT, Massachusetts, 205-219. http://kryten.mm.rpi.edu/Divine-Command_Roboethics_Bringsjord_Taylor.pdf

Redukcionizem, samorazumevanje in učinki zankanja v kognitivni znanosti

Reductionism, Self-Understanding, and Looping Effects in Cognitive Science

Ema Demšar
Center za kognitivno znanost
Pedagoška fakulteta
Univerza v Ljubljani
Slovenija
ema.demsar@pef.uni-lj.si

IZVLEČEK

V pričujočem prispevku raziščem odnos med nevroznanstvenimi pojmovanji človekove duševnosti, njihovo predstavitvijo v javnosti in načini, na katere se vpletajo v samorazumevanje ter konkretni vsakdan posameznikov v sodobni družbi. Po kratki omembi teoretske delitve na znanstveno in manifestno podobo človeka ponudim pregled izbranih empiričnih raziskav s področja širjenja nevroznanstvenih idej v medijih in njihove integracije v vsakdanjem (samo)razumevanju posameznikov. S pomočjo koncepta zankanja človeških vrst pokažem na kompleksnost odnosa med opisovanjem duševnih pojavov v nevroznanosti in tega, kako se ti pojavi kažejo v sodobnem življenjskem svetu. Izpostavim, da se v vsakdanje pojmovanje duševnosti ne vključijo nujno tisti nevroznanstveni koncepti, ki so najbolje podprti z raziskavami, ampak tisti, ki jih je mogoče integrirati z obstoječimi družbeno-kulturnimi motivacijami in okvirji prepričanj.

Ključne besede

Kognitivna znanost, nevroredukcionizem, samorazumevanje, učinki zankanja

ABSTRACT

In this contribution, I explore the relationship between neuroscientific conceptions of the human mind and mental phenomena, their presentation in the public sphere, and the ways in which they become involved in the self-understanding and day-to-day lives of individuals in contemporary society. After briefly touching on the theoretical distinction between the (neuro)scientific and the manifest image of the human being, I offer an overview of selected empirical research on the dissemination of neuroscientific conceptions of the mind in the media and their integration into the everyday self-understanding of individuals. With the help of the concept of the looping effect of human kinds, I point to the complexity of the relationship between describing mental phenomena in neuroscience and how these phenomena are manifested in the modern lifeworld. I emphasize that everyday conceptions of the mind do not necessarily include those neuroscientific concepts that are best supported by research, but those that can be best integrated with existent socio-cultural frameworks of beliefs and motivations.

Keywords

Cognitive science, neuroreductionism, self-understanding, looping effects

1. Uvod

Razvoj nevroznanosti v zadnjih treh desetletjih je ob znanstvenih odkritjih in tehnološkem napredku pripeljal tudi do porasta prisotnosti nevroznanstvenih idej v javnem prostoru. Če je bil leta 1990 vodilni razlog za razglasitev *desetletja možganov* (ang. *decade of the brain*), kot je George Bush v sodelovanju z znanstvenimi ustanovami poimenoval devetdeseta leta prejšnjega stoletja, »povečati zavedanje javnosti o koristih, ki si jih lahko obeta od raziskovanja možganov«[1], se danes zdi, da je javnost dodobra ozaveščena. Skupaj z razširitvijo dometa nevroznanstvenih raziskav na polje raziskovanja čustev ter socialnega in kulturnega sveta postajajo nevroznanstvene informacije relevantne na mnogih področjih človekovega delovanja, ki ležijo onkraj prvotnih aplikacij v znanosti in zdravstvu. Nevroznanstvena pojmovanja duševnosti in duševnih pojavov ter govor o možganih (ang. *brain talk*) [1] so se iz laboratorijev in klinik razširili tako v medije [2] in vsakdanjo pogovorno rabo [3] kot tudi v prostore šol, ministrstev in gospodarskih zbornic [4].

V promocijskih materialih ob zagonu evropskega nevroznanstvenega *Projekta človeški možgani* (ang. *The Human Brain Project*) najdemo trditev, da bi, če projektu uspe zastavljeni cilj – tj. uspešno simulirati možgane – to »globoko vplivalo na naša najbolj temeljna prepričanja – še posebej na naše razumevanje sebstva, svobodne volje in osebne odgovornosti, načina, na katerega sami sebe razumemo kot osebe, ki so osebno odgovorne za svoja dejanja«[2]. Do katere mere se je z vzponom nevroznanosti naše samorazumevanje – in razumevanje tega, kaj pomeni biti človek – res spremenilo pod vplivom širjenja nevroznanstvenih pojmovanj duševnosti? Po številnih teoretskim obravnavah je to vprašanje v zadnjih letih tudi nekaj empiričnih raziskav širjenja nevroznanstvenih informacij v medijih ter njihovega vključevanja v življenjske svetove različnih članov družbe.

V tem prispevku se bom pomočjo pregleda izbranega nabora empiričnih študij vprašala, na kakšne načine se v sodobnem svetu nevroznanstvene informacije prepletajo z vsakdanjim razumevanjem človekove duševnosti in skušala tako empirično

[1] Pridobljeno s http://www.loc.gov/loc/brain/proclaim.html (september 2019), lastni prevod.

[2] Pridobljeno s http://www.humanbrainproject.eu/ethics.html (december 2014), lastni prevod.

osvetliti staro filozofsko vprašanje o povezavi med znanstveno in manifestno podobo človeka.

2. Manifestna in (nevro)znanstvena podoba človeka

Razkol med pojmovanjem duševnosti v vsakdanjem življenju in znanstvenimi razlagami duševnih pojavov seveda ni le stvar sodobne kognitivne znanosti. Na perečo vrzel med t. i. življenjskim svetom vsakdanjega izkustva in svetom, kot ga prikazujejo znanstvene discipline, je že v prvi polovici prejšnjega stoletja opozarjal Edmund Husserl [5], najbolj izrecno pa je na razkorak med življenjskim in znanstvenim gledišcem znotraj razumevanja duševnosti pokazal ameriški filozof Wilfrid Sellars v svojem razlikovanju med t. i. manifestno in znanstveno podobo človeka v svetu [6]. *Manifestna podoba* predstavlja pojmovni okvir, znotraj katerega ljudje dojemamo sebe in druge kot osebe, ki živijo in delujejo v vsakdanjem človeškem svetu pomenov, norm, namer in osebne odgovornosti. V nasprotju s tem *znanstvena podoba* človeka obravnava kot kompleksen fizični sistem – v primeru nevroredukcionističnih pogledov, s katerim se bom ukvarjala v pričujočem prispevku, kot »nič več kot skupek živčnih celic« [7: str. 2, lastni prevod].

Kljub zaupanju v ontološko primarnost znanstvene podobe je Sellars razkorak med obema podobama prepoznal kot temeljni filozofski problem sodobnega časa. Ker se koncepti, potrebni za vsakdanje razumevanje sebe in drugih kot oseb, neizogibno izgubijo, kadar zvedemo duševnost na objektivistične znanstvene razlage, je vztrajal, da je manifestna podoba nujno potrebna za razumevanje normativnih in pomenskih vidikov človekovega življenja. Po Sellarsu zato celovito razumevanje človeka v svetu zahteva zedinjen pogled (t. i. »stereoskopski vid«), ki zmore obenem zaobjeti tako znanstveno kot manifestno podobo.

V zadnjih desetletjih je skupaj s porastom prisotnosti nevroznanstvenih idej v javnosti naraslo tudi število analiz, ki kažejo, da se (nevro)znanstvena podoba človeka vse bolj vpleta v pojmovni okvir našega vsakdanjega razumevanja lastne duševnosti in duševnosti drugih. Danes številni teoretiki opozarjajo, da se kot posledica razvoja ter popularizacije nevroznanosti in povečanega vpliva nevrotehnologije in psihofarmakologije vsakdanje razumevanje duševnosti v sodobni (predvsem zahodni) družbi vedno tesneje tke okrog konstruktov, kakršni so »nevrokemično sebstvo« (ang. *neurochemical self*) [8], »možganski subjekt« (ang. *cerebral subject*) [9] in »možganstvo« (ang. *brainhood*) [3].

Redukcija duševnih pojavov na njihove nevrobiološke korelate, ki izkustvo obravnava kot epifenenomen brez vzročne vloge v delovanju duševnosti, socialni svet pa kot neodvisni nabor zunanjih dražljajev, ima lahko pomembne posledice za človeško samorazumevanje in različne vidike urejanja družbe [4]. Petdeset let po tem, ko je Sellars izpostavil razkorak med znanstveno in manifestno podobo, se tako zdi njegov klic po zedinjenem pogledu še posebej relevanten. Po drugi strani pa sodobna vprašanja o povezavi med nevroznanostjo, družbo in samorazumevanjem posameznikov v družbi ponujajo edinstveno priložnost za empirično osvetlitev njegove in drugih filozofskih razprav o razkoraku med obema pogledoma.

3. Nevroznanstvene razlage duševnosti v medijih

Ob bliskovitem razvoju tehnologije za slikanje možganov s funkcijsko magnetno resonanco so nevroznanstvene raziskave človeških možganov s prvotnega fokusa na senzorimotorične in kognitivne procese kmalu posegle na področje pojavov, ki tradicionalno spadajo v domeno družboslovja in humanistike [10]. Danes med predmeti nevroznanstvenih raziskav neredko najdemo fenomene, ki so najbolj intimno zvezani z našim samorazumevanjem in razumevanjem vprašanja, kaj pomeni biti človek: od ljubezni, umetnosti in religije do politike in prava. Kot so v pregledu britanskih časopisnih člankov, izdanih med letoma 2000 in 2010, pokazali O'Connor, Rees in Joffe [2], je pričetek stoletja prinesel dramatičen porast poročanja o nevroznanstvenih raziskavah v medijih. Čeprav uporaba znanstvenih konceptov za podkrepitev vsakdanjih trditev o duševnosti ni novost [11], se zdi, da so z novimi nevroslikovnimi metodami podkrepljene nevroznanstvene razlage duševnosti še posebej prepričljive. V zgodnjih raziskavah vpliva prisotnosti nevroznanstvenih informacij v besedilu so bralci argumente v splošnem presojali kot bolj kredibilne, kadar so jih spremljali (za samo vsebino argumenta nerelevantni) nevroznanstveno izrazje in slike [12, 13]. Ta retorična moč nevroznanosti se s pridom izrablja v medijih, v katerih informacije o možganih (še posebej nevroslikovni material) pogosto služijo kot podkrepitev v člankih podanih razlag – tudi, kadar trditvam ne pridajo nobene dejanske razlagalne vrednosti.

V medijski analizi iz leta 2005 so Racine, Bar-Ilan in Illes [14] prepoznali tri glavne načine, na katere je nevroznanost pogosto izrabljena kot retorično orodje: uporabo nevroznanstvenih informacij za utemeljevanje resničnosti ali objektivnosti raziskovanega pojava (*nevrorealizem*), interpretiranje možganov kot bistva osebe, pri čemer pojem »možgani« navadno zamenja koncepte, kot so »duševnost«, »jaz« ali »sebstvo« (*nevroesencializem*), ter uporabo nevroznanstvenih študij za promocijo in podporo političnih ali osebnih ciljev (*nevropolitika*). Kasnejše analize [2] so dodatno identificirale še naraščajoči trend prikaza možganov kot vira za samoizboljšavo in optimizacijo »možganskih« oz. psiholoških funkcij, pa tudi trend posluževanja nevroznanstvenih informacij za poudarjanje nevrobioloških variacij med različnimi demografskimi ali diagnostičnimi skupinami (npr. med spoloma in spolnimi usmerjenostmi, med kriminalno in nekriminalno ali klinično in neklinično populacijo). Slednja strategija, ki (povezana z zgoraj omenjenima nevrorealizmom in nevroesencializmom) skuša s sklicevanjem na nevrobiološke razlike med skupinami razložiti razlike v njihovih vedenjskih in psiholoških značilnostih, najpogosteje temelji na (navadno implicitni) biologizaciji družbenih kategorij in je tako pogosto izrabljena, da na novo interpretira družbeno oblikovane pojave kot posledico »naravnega reda« [15].

4. Nevroznanstvene ideje v samorazumevanju posameznikov

Trendi prikazovanja nevroznanosti v medijih pa ne odražajo nujno načinov, na katere se nevroznanstvene informacije vpletajo v dejanski vsakdan njihovega občinstva. V redkih raziskavah s področja uporabe nevroznanstvenih informacij v kontekstu razumevanja duševnosti pri dejanskih posameznikih se je pokazalo, da je zanimanje splošne javnosti za nevroznanstvene podatke najverjetneje skromnejše, kot bi to predvideli na podlagi »nevromanije« v medijih [16]. Medtem ko posamezniki koncept možganov pogosto dojemajo kot relevantnega v kontekstu abstraktnih razprav, v konkretnem vsakdanu povprečnega člana družbe navadno ne zavzema posebej pomembne vloge za samorazumevanje in razumevanje duševnosti [17].

V primerjavi z omejenim prevzemanjem nevroznanstvenih idej v splošni javnosti nosijo nevroznanstveni pojmi opazno večji pomen v skupnostih, v katerih se jih lahko uporabi kot orodje za samorazumevanje in socialno prezentacijo. Vpletanje nevrobioloških konceptov in nevroredukcionističnih pogledov v razlago svojega stanja in gradnjo osebne identitete je še posebej pogosto v klinični populaciji posameznikov s psihiatrično diagnozo. V skladu z zgoraj navedenimi trendi medijske uporabe nevrorealizma in nevroesencializma nevroznanstvene – predvsem nevroslikovni material – informacije pogosto služijo kot prepričljivo orodje za samointerpretacijo: s svojim potencialom za reifikacijo duševne bolezni v fizičnem substratu možganov za mnoge posameznike predstavljajo »dokaz za biološki obstoj [njihove] duševne bolezni« [18, str. 18]. Kot kaže nabor raziskav s pacienti z razpoloženjskimi motnjami, pa se posamezniki tega »dokaza« poslužujejo za različne vrste (samo)interpretacije. Medtem ko za nekatere predstavlja orodje za opolnomočenje, manjšanje stigme, legitimizacijo njihovega stanja ali prelaganje sebi pripisane osebne odgovornosti zanj, lahko isti nevroznanstveni koncepti, ideje in/ali razlage pri drugem posamezniku ali v drugem kontekstu vodijo do nasprotnih učinkov, npr. v resignacijo spričo svoje diagnoze, povečanje stigme ali zvišanje sebi pripisane osebne odgovornosti za svoje stanje [18–22]. Tako se zdi, da za klinično populacijo nevroznanstvene informacije ne *določajo* samointerpretacije, temveč služijo kot potencialni *material* zanjo – material, ki glede na dani kontekst omogoča različne, včasih celo nasprotujoče si načine narativnega uokvirjanja diagnoze v širši kontekst pacientovega življenja in osebne identitete.

5. Učinki zankanja človeških vrst

Spremembe v razumevanju duševnosti – bodisi v konkretnih življenjskih svetovih posameznikov bodisi na nivoju medijskih reprezentacij – se ne dogajajo v vakuumu. Kot izpostavljata Nikolas Rose and Joelle Abi-Rached [4], vprašanje, kaj pomeni biti človek, ni le predmet filozofskih razprav, temveč nosi pomembne praktične posledice. Pojmovanje duševnosti in duševnih pojavov v družbi igra ključno vlogo pri tem, kako kot družba načrtujemo, vodimo in urejamo svoje izobraževalne, pravne in kazenske sisteme, svoje socialne in gospodarske politike, zdravstvo in psihiatrijo, pa tudi svoje estetske in etične okvirje – po drugi strani pa vsi ti sistemi in okvirji »upravljajo« in »vodijo« nas same.

Pomembno je poudariti tudi, da oblikovanje pojmovanj duševnosti ni le enosmeren proces iz nevroznanstvenega laboratorija v družbo. Medtem ko rezultati nevroznanstvenih raziskav informirajo družbeno in osebno razumevanje duševnosti, je nevroznanost (in širše kognitivna znanost) kot socialna aktivnost tudi sama umeščena v svoje družbeno in politično okolje, družbeno pojmovanje duševnih pojavov pa se – še posebej v primeru preučevanja človeških možganov in duševnosti – vpleta v raziskovalni proces od izbire raziskovalnega vprašanja, udeležencev in metod za pridobivanje ter analizo podatkov do interpretacije pridobljenih rezultatov.

Zaradi posledic, ki jih imajo nevroznanstvene ideje na vsakodnevno pojmovanje duševnosti v družbi, lahko širjenje rezultatov raziskav v javnem prostor posredno vzvratno vpliva na proces raziskovanja v nevroznanosti, kjer z izbiro uporabljene metodologije nevroznanost *sooblikuje* – in ne »le« preučuje – prav tiste pojave, ki jih skuša razložiti [19].

Filozof Ian Hacking proces dvosmernega vzajemnega sovplivanja med opisovanjem in klasifikacijo duševnih pojavov v znanosti ter njihovim obstojem v družbi in življenjskem svetu posameznikov zajame v konceptu *učinka zankanja človeških vrst* [23]. Med človeške vrste uvršča pojave, ki so – za razliko od t. i. *naravnih vrst* – po definiciji umeščeni v določeno družbeno in konceptualno okolje. Zaradi te umeščenosti človeške vrste, med katere spadajo mnogi psihološki konstrukti [24], *interagirajo* z opisi in klasifikacijami, ki so jim pripisani. Točneje, opisovanje in klasificiranje človeških vrst lahko vpliva na način, na katerega se te vrste manifestirajo v vsakdanu.

Razširitev določenega pojmovanja duševnega pojava (npr. pojmovanja duševne bolezni, kakršna je depresija, kot »bolezni možganov« raje kot »bolezni osebe« [25]) v znanosti in medijih vpliva na to, kako se ta pojav razume in obravnava v družbi. Posameznikom, za katere je pojav relevanten (v tem primeru ljudem, ki trpijo za depresijo), so kot posledica spremenjenega razumevanja na voljo vsaj delno drugačne intervencije in možnosti delovanja (npr. spodbuda k psihofarmakološkem zdravljenju raje kot k integrativni psihoterapiji) ter drugačen pojmovni okvir za samorazumevanje in konstrukcijo osebne identitete. To vodi do sprememb v njihovem doživljanju in dejanskem vedenju – sprememb, ki dalje vplivajo na pojmovanje danega pojava v družbi (vključno z načinom, na katerega se znanstveno raziskuje). Tako znanstvena pojmovanja duševnih pojavov, kot je depresija, pomembno sooblikujejo način, na katerega razumemo, doživljamo in živimo depresijo v kontekstu vsakdanjega življenja – istočasno pa način, na katerega se depresija živi in manifestira v vsakdanu, podpira naše znanstveno pojmovanje tega pojava.

Za razliko od analiz, ki svarijo pred enoznačnim določanjem vsakodnevnega pojmovanja duševnosti z nevroznanstvenimi (in bolj specifično nevroredukcionističnimi) pogledi, se zdi koncept zankanja človeških vrst primernejši za razumevanje raznolikih in dinamičnih načinov prepletanja (nevro)znanstvene in manifestne podobe človeka v sodobni družbi. Kot opozarja antropologinja Emily Martin [20], v izgradnji konceptualnih okvirov za vsakdanje razumevanje duševnosti ne prevladajo nujno modeli, ki so najbolje podprti z rezultati nevroznanstvenih raziskav: ohranijo se tiste razlage, ki jih je možno najbolj smiselno integrirati v trenutno kulturno in družbeno-politično okolje. Na področju psihiatrije, na primer, med razlogi za privlačnost nevroredukctionističnih razlag – tudi takih, ki jim primanjkuje empirične podpore, kakršna je na primer monoaminska hipoteza depresije – najdemo vpliv interesov in finančnih investicij psihofarmacevtske industrije, pa tudi zmožnost nevrobiologije, da preusmeri pozornost s politično spornih družbenih in ekonomskih dejavnikov, ki doprinašajo k nastanku duševnih bolezni [26], po drugi strani pa nevroredukcionistične razlage s prenosom vzročnosti z delovanja *osebe* na delovanje *možganov* mnogim posameznikom s psihiatričnimi diagnozami omogočajo zmanjšanje občutka, da sami povzročajo svoje trpljenje. Uporaba nevroredukcionističnih konceptov je v medijih pogosto promovirana z odkrito aktivističnimi cilji, kot sta zmanjšanje stigme in spodbuda oblikovanja pozitivne identitete prizadetih skupnosti; najopaznejša primera najdemo pri »gibanju za nevrodiverziteto« v avtistični skupnosti [3, 22] ter pri otrocih z motnjo pozornosti in hiperaktivnosti [1].

Morda še pogosteje kot v spreminjanje vsakdanjega razumevanja duševnosti pa biologizacija človeških vrst vodi v utrjevanje že obstoječih prepričanj in stereotipov tako v medijskih reprezentacijah duševnosti kot tudi pri samorazumevanju posameznikov. Vidal and Ortega [22, str. 17] tako govorita o

»soobstoju ontologij« in »sobivanju konceptov sebstva«, s pomočjo katerih ljudje v svojem samorazumevanju in socialni prezentaciji pogosto prehajamo med mnogimi registri govora o duševnosti. Ni neobičajno, denimo, da se ista oseba interpretira s pomočjo sklicevanja na pojme, ki izhajajo iz različnih – celo nasprotujočih si – sklopov pojmovanj duševnosti. Medtem ko »brain talk« včasih zares implicira redukcijo duševnosti na »skupek nevronov« osrednjega predstavnika nevroredukcionizma Francisa Cricka [7], koncept možganov v medijih [27] in vsakdanji rabi [3, 11] pogosto služi kot metafora za širok razpon različnih pojmovanj ter teorij – od nevrokemičnih do psihoanalitičnih – o duševnih pojavih in sebstvu, pri čimer navadno ohrani psihološko globino, ki je bila pred vzponom nevroznanosti pripisana duševnosti.

6. Zaključek

Na podlagi pregleda izbranih empiričnih raziskav lahko sklenemo, da se dandanes (nevro)znanstvena podoba v vsakodnevno pojmovanje duševnosti vpleta na fleksibilne in raznolike načine. Kljub pričakovanjem, da bo nevroznanost vsak čas pripeljala do radikalne revolucije v pojmovanju duševnosti [28], se zdi, da nova nevroznanstvena spoznanja niso zmanjšala zaupanja v duševno in psihološko domeno, znotraj katere učinkujejo [4]. Integracija izbranih nevrobioloških konceptov v vsakodnevno razumevanje duševnosti je poleg prepričljivosti znanstvenih rezultatov odvisna od različnih motivacij za redukcionistični pogled – področja, na katerih se nevroredukcionistične ideje »zakoreninijo«, pa odražajo predvsem njihovo skladnost z že obstoječimi prepričanji o duševnosti in duševnih pojavih [3, 4, 16–20].

7. Reference

[1] Singh, I. 2013. Brain talk: power and negotiation in children's discourse about self, brain and behaviour. *Sociology of Health and Illness*, 35, 6, 813–827. DOI= https://doi.org/10.1111/j.1467-9566.2012.01531.x.

[2] O'Connor, C., Rees, G., and Joffe, H. 2012. Neuroscience in the public sphere. *Neuron*. 74, 2, 220–226. DOI= https://doi.org/10.1016/j.neuron.2012.04.004.

[3] Vidal, F. 2009. Brainhood, anthropological figure of modernity. *History of the Human Sciences*, 22, 1, 5–36. DOI= http://dx.doi.org/10.1177/0952695108099133.

[4] Rose, N., and Abi-Rached, J. 2014. Governing through the brain: Neuropolitics, neuroscience and subjectivity. *The Cambridge Journal of Anthropology*, 32, 1, 3–23. DOI= https://doi.org/10.3167/ca.2014.320102.

[5] Husserl, E. 1970. *The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy* (D. Carr, Trans.). Northwestern University Press, Evanston, IL.

[6] Sellars, W. 1963. *Science, Perception and Reality*. Humanities Pres, New York.

[7] Crick, F. 1994. *The Astonishing Hypothesis.* Charles Scribner's Sons, New York.

[8] Rose, N. 2003. Neurochemical selves. *Society*, 4, 1, 46–59. DOI= https://doi.org/10.1007/BF02688204.

[9] Ortega, F. 2009. The cerebral subject and the challenge of neurodiversity. *BioSocieties*, 4, 4, 425–445. DOI= https://doi.org/10.1017/s1745855209990287.

[10] Illes, J., Kirschen, M. P., and Gabrieli, J. D. 2003. From neuroimaging to neuroethics. *Nature Neuroscience*, 6, 3 (2003), 205–205. DOI= http://dx.doi.org/10.1038/nn0303-205.

[11] Rose, N. 2001. The Politics of Life Itself. *Theory, Culture & Society*, 18, 6, 1–30.

[12] Weisberg, D. S., Keil, F. C., Goodstein, J., Rawson, E., and Gray, J. R. 2008. The Seductive Allure of Neuroscience Explanations. *Journal of Cognitive Neuroscience*, 20, 3, 470–477. DOI= https://doi.org/10.1162/jocn.2008.20.3.470.

[13] Fernandez-Duque, D., Evans, J., Christian, C., and Hodges, S. D. 2015. Superfluous neuroscience information makes explanations of psychological phenomena more appealing. *Journal of Cognitive Neuroscience*, 27, 5, 926–944. DOI= https://doi.org/10.1162/jocn_a_00750.

[14] Racine, E., Bar-Ilan, O., and Illes, J. 2005. fMRI in the public eye. *Nature Reviews Neuroscience*, 6, 2, 159–164. https://doi.org/10.1038/nrn1609.

[15] Dupré, J. (2001). *Human Nature and the Limits of Science.* Oxford University Press, Oxford.

[16] O'Connor, C., and Joffe, H. 2013. How has neuroscience affected lay understandings of personhood? A review of the evidence. *Public Understanding of Science*, 22, 3, 254–268. DOI= https://doi.org/10.1177/0963662513476812.

[17] Pickersgill, M., Cunningham-Burley, S., and Martin, P. 2011. Constituting neurologic subjects: Neuroscience, subjectivity and the mundane significance of the brain. *Subjectivity*, 4, 3, 346–365. DOI= https://doi.org/10.1057/sub.2011.10.

[18] Dumit J. 2003. Is it me or my brain? Depression and neuroscientific facts. *J. Med. Humanit*. 24, 35–47 DOI= http://dx.doi.org/10.1023/A:1021353631347.

[19] Dumit, J. 2004. *Picturing personhood: Brain scans and biomedical identity.* Princeton University Press, Princeton NJ.

[20] Martin, E. 2010. Self-making and the brain. *Subjectivity* 3, 4 (Aug. 2010), 366–381. DOI= http://dx.doi.org/10.1057/sub.2010.23.

[21] Slaby, J. 2010. Steps towards a critical neuroscience. *Phenomenology and the Cognitive Sciences*, 9, 3, 397–416. DOI= https://doi.org/10.1007/s11097-010-9170-2.

[22] Vidal, F., and Ortega, F. 2011. Approaching the neurocultural spectrum: An introduction. In *Neurocultures: Glimpses into an expanding universe*. Lang, 7–27.

[23] Hacking, I. 1995. The looping effects of human kinds. In *Causal Cognition* (Nov. 1996), 351–383.

[24] Brinkmann, S. 2005. Human kinds and looping effects in psychology: Foucauldian and hermeneutic perspectives. *Theory & Psychology*, 15, 6, 769–791. DOI= https://doi.org/10.1177/0959354305059332.

[25] Fuchs, T. 2012. Are Mental Illnesses Diseases of the Brain? In *Critical Neuroscience*, 331–344. Wiley-Blackwell, Chichester.

[26] Kirmayer, L. J., and Gold, I. 2012. Re-socializing psychiatry. In *Critical neuroscience*, 305–330. Wiley-Blackwell, Chichester.

[27] Whiteley, L. 2012. Resisting the revelatory scanner? Critical engagements with fMRI in popular media. *BioSocieties*, 7, 3, 245-272. DOI= https://doi.org/10.1057/biosoc.2012.21.

[28] Lynch, Z. 2009. *The Neuro Revolution.* St. Martin's Press, New York.

Modelling natural selection to understand evolution of perceptual veridicality and its reaction to sensorimotor embodiment

Tine Kolenik
Jožef Stefan Institute & Jožef Stefan
International Postgraduate School
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3807
tine.kolenik@ijs.si

ABSTRACT
The relationship between mind and world has always been one of the focal interests of cognitive science. Perception has been identified as one of the main sources of knowledge about the world and therefore a prime research interest. Evolutionary scientists claim that natural selection optimizes perception so that it accurately mirrors the outside world. In opposition, the interface theory of perception proposes that perception is a non-veridical interface between an organism and the outside world, evolutionarily fitted to the organism's fitness and not the objective truth. It has been studied using genetic algorithms (GAs), which show that non-veridical perception offers more survival value to the modelled organism than veridical perception. However, the theory is based on cognitivist presuppositions about the mind, claiming that perception does not require action. We successfully replicated the GA model, then replaced cognitivist presuppositions with embodied-enactivist presuppositions, coupling action and perception by adding a sensorimotor loop. The sensorimotor loop bootstraps evolution, with organisms needing less information to perform better due to knowing how to perceive by taking appropriate actions. We also perform additional experiments to further corroborate our claims.

Keywords
Cognitivism, enactivism, evolution, genetic algorithms, interface theory of perception.

1. INTRODUCTION
Perceptions have evolved not to describe the objective world, but to help us survive. In a way, they are similar to a computer desktop, which shows its elements, like icons, as to make them easily manipulatable, but 'hides the truth' behind them, like the underlying electrical current. This is the main idea of the interface theory of perception (ITP) [1].

Hoffman et al. [1] claim that perceptions are not isomorphic – "a structure-preserving relation between the physical-causal make-up of the system and the formal structure of the computational model supposedly instantiated by the system" [2, p. 7] – to the objective world, but to the evolutionary fitness of the perceiving organism. ITP therefore follows a more general upheaval in cognitive science (predictive coding [3], enactive approaches [4]) that goes against the idea that perception generates "a fully spatial virtual-reality replica of the external world in an internal representation." [5, p. 375]. Hoffman et al. use, among other methods, genetic algorithms (GAs) to back up their theory [6]. Their model generates a population of artificial organisms that can perceive and act, and evolves them. After a number of generations, the organisms that survive and reproduce do not perceive the objective world isomorphically – rather, they perceive it according to their internal needs, isomorphic to their payoff function.

In our work, we replicate their GA model. Hoffman et al. make a claim that perceptual experience does not require motor movement [1, p. 1497]. We believe that is not true, following enactive approaches to sensorimotor cognition [7], and make our own GA model. In it, we replace cognitivist presuppositions on sensomotorics with embodied-enactivist ones by adding a sensorimotor loop. This also serves to offer further evidence for ITP's idea.

2. REPLICATION
Hoffman et al.'s cognitivist model (CM) is based on Mitchell's 'Robby, the Soda-Can-Collecting Robot' [8]. Robby is an agent that forages soda cans scattered on a grid (Figure 1). It can make a move in a Von Neumann neighborhood (non-diagonally adjacent cells), which it perceives, as well as try to pick up a soda can. It gets points if there is a soda can in the cell it stands on. It loses points if there is no soda can or if it bumps into a wall surrounding the grid. The GA model generates many such grids with many Robbies, who start out with very bad strategies for foraging. Through evolution, where Robbies with better strategies are selected for DNA crossover, Robbies in the final generation become masters of their craft. Their DNA is composed of situation-move pairs, where the situation part describes a possible configuration of soda cans in a Von Neumann neighborhood, and the move part describes which move to make when Robby is in that situation.

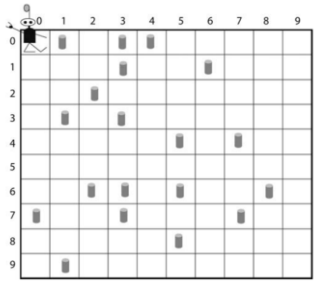


Figure 1: Robby and its world [from 6, p. 131].

Hoffman et al. modify Mitchell's model in a number of ways to be able to investigate ITP. They add a perceptual DNA (pDNA) to Robbies alongside their foraging DNA (fDNA) to evolve as well. The pDNA determines how Robbies see the cells in their Von Neumann neighborhood. They either see them colored in red or in green, depending on the number of soda cans in the perceived cell and their pDNA. As implied, Hoffman et al. also changed the number of possible soda cans in a cell from up to 1 to up to 10. The points Robbies get from picking up soda cans are modified as well – the payoff function is Gaussian, Robbies get (0,1,3,6,9,10,9,6,3,1,0) points for (0,1,2,3,4,5,6,7,8,9,10) cans, respectively (see Figure 2). Each gene in the pDNA represents one amount of soda cans, connecting it with one of the two colors.

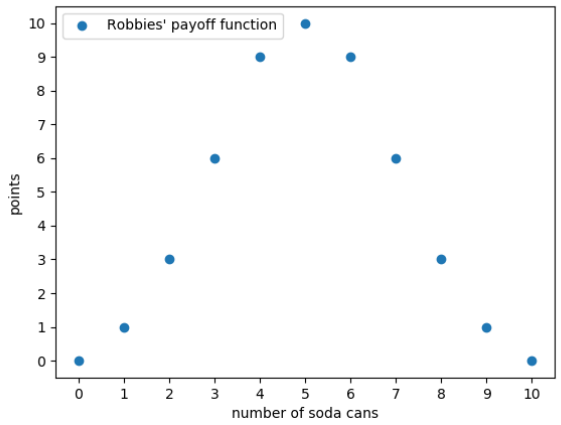


Figure 2: Robbies' Gaussian payoff function for foraging soda cans.

Robbies evolve similarly as in Mitchell's model – they start with bad strategies and end with good ones. What is of interest is how their pDNA evolves during this time – the question is whether the perception is isomorphic or non-isomorphic to the outside world. If the pDNA were to evolve to be isomorphic, it would look like the top genome in Figure 3, which makes colors organize to reflect the lower and the higher amounts of soda cans. If it were to evolve to be non-isomorphic, it would look like the bottom genome in Figure 3, reflecting Robbies' fitness function. It is the latter that does evolve, making Robbies not see the world isomorphic to the outside world, but in a way that helps them survive – the number of soda cans that brings them the most points are of one color, the number that brings them the least points are of another color.

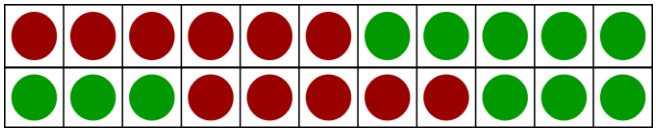


Figure 3: Isomorphic (top) and non-isomorphic (bottom) perceptual DNA.

3. EMBODIED-ENACTIVE MODEL

We look at ITP from an embodied-enactive perspective [7], especially since Hoffman et al. claim that perception is possible without action. Therefore, we add a sensorimotor loop to Robbies. Our model's (EAM) modifications are the following: previously able to see the Von Neumann neighborhood, now Robbies only see the cell they are in and the cell they are looking at. The latter implies another modification – Robbies first have to act to perceive. They have to turn towards a certain direction to see the cell in that direction. Robby therefore has the following 'loop of life':

1. Depending on where Robby is looking at, perceive the cell's color.
2. Make a move depending on what Robby sees in the direction it is looking at and the cell it is standing on.
3. Decide which cell to turn to, which will be perceived in step 1 of the process' reiteration.

The fDNA is modified to include turn-situation-move triplets, which are then evolved instead of only situation-move pairs as in

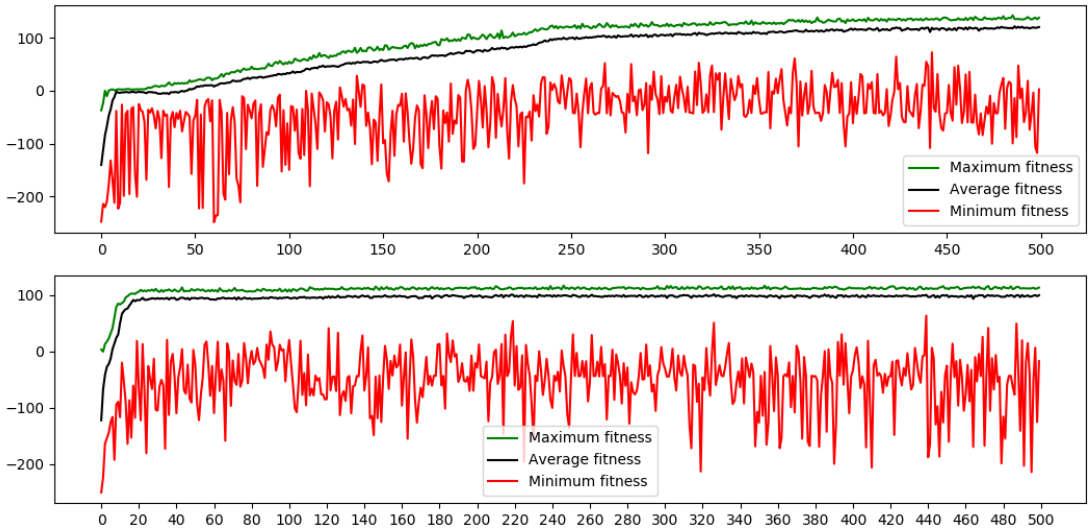


Figure 4: Robbies' foraging skills evolution in CM (top) and EAM (bottom).

CM. Figure 4 shows the results of how Robbies and their fitness (number of points on y-axis) evolve (time on x-axis) with CM on the top and EAM on the bottom. EAM's Robbies' pDNA evolves the same as in CM.

4. ADDITIONAL EXPERIMENTS

Four additional experiments were made with CM and EAM to further examine legitimacy of non-isomorphic perception prevailing over isomorphic perception. Robbies were implemented with pDNA coding the mapping from the external world to colors that was constant, unchanged neither by crossover nor by mutation. Four experiments were run:

1. CM was implemented with a fixed isomorphic perceptual strategy.
2. CM model was implemented with a fixed non-isomorphic perceptual strategy.
3. EAM was implemented with a fixed isomorphic perceptual strategy.
4. EAM was implemented with a fixed non-isomorphic perceptual strategy.

Figures 5, 6, 7 and 8 show graphs for CM with a fixed isomorphic perceptual strategy, CM with a fixed non-isomorphic perceptual strategy, EAM with a fixed isomorphic perceptual strategy and EAM with a fixed non-isomorphic perceptual strategy, respectively.

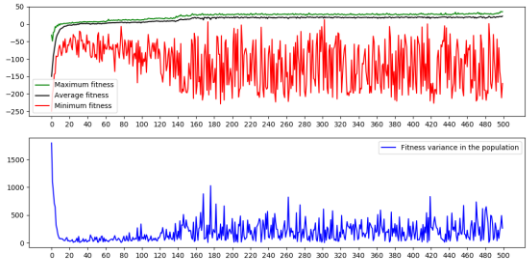


Figure 5: CM with a fixed isomorphic perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

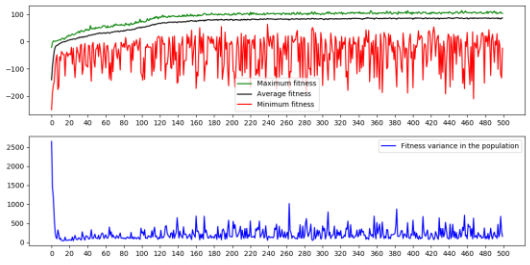


Figure 6: CM with a fixed non-isomorphic perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

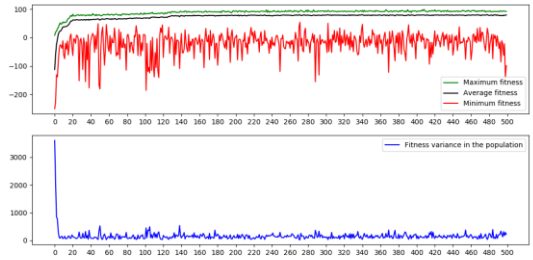


Figure 7: EAM with a fixed isomorphic perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

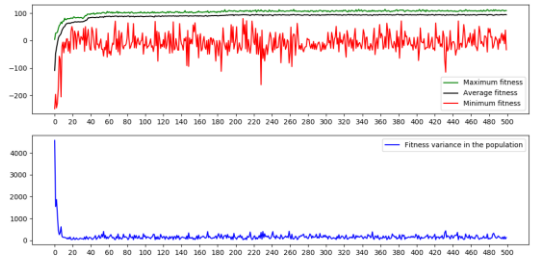


Figure 8: EAM with a fixed non-isomorphic perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

Experiments mostly yielded nothing out of the usual. Both models with non-isomorphic perceptual strategies scored similarly between each other as well as to the original models without fixed, but evolving perceptual strategies. The slope of CM's two graphs compared to EAM's are again to be expected – the same happened in the models with evolving strategies. The same goes for variance. What is unexpected is that Robbies with isomorphic perceptual strategies in EAM score a lot higher than Robbies with the same perceptual strategy in CM. This might be again due to the varying variance and higher scoring individuals in EAM, where the sensorimotor loop works as an optimizer.

Further experiments therefore yielded results that were expected, and showed that the fitness-based, non-isomorphic perceptual strategy makes Robbies more successful in picking up soda cans and navigating the modelled world.

5. DISCUSSION AND CONCLUSIONS

CM and EAM both evolve perceptions that are not isomorphic to the objective world, but rather to the perceiving organism's needs. However, they diverge in how long it takes for Robbies to become master foragers. EAM implements active perception [9], which bootstraps evolution and optimizes the best foraging strategy discovery process. This means that actively choosing which (and less) information to take in beats more ('free') information which needs to be processed in CM. In our future work, we want to make Robbies more 'enactively' autonomous [10], meaning that there would be less designer-fixed agent architectures and more learning through non-deterministic dynamic interactions. We also want their fitness function more dependent on historical interactions [11]. Lastly, we want to conceptualize the role of such modelling in researching how presuppositions of different

cognitive science paradigms influence our understanding of cognition [12].

6. ACKNOWLEDGEMENTS

I thank Urban Kordeš for his mentorship of the Master's thesis, which served as the foundation for this work.

7. REFERENCES

[1] Hoffman, D. D., Singh, M. and Prakash, C. 2015. The interface theory of perception. *Psychonomic Bulletin & Review*. 22 (2015), 1480–1506. DOI= https://doi.org/10.3758/s13423-015-0890-8.

[2] Haselager, de Groot, & van Rappard, 2003, p. 7

[3] Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. 36 (2013), 181-204. DOI= https://doi.org/10.1017/s0140525x12000477.

[4] Ward, D., Silverman, D., and Villalobos, M. 2017. Introduction: The varieties of enactivism. Topoi, 36, 3 (2017), 365–375. DOI= https://doi.org/10.1007/s11245-017-9484-6.

[5] Lehar, S. 2003. Gestalt isomorphism and the primacy of subjective conscious experience: a Gestalt Bubble model. *Behavioral and Brain Sciences*, 26, 4 (2003), 375–444. DOI= https://doi.org/10.1017/s0140525x03000098.

[6] Mitchell, M. 1999. *An introduction to genetic algorithms*. The MIT Press, Cambridge, MA.

[7] Di Paolo, E., Buhrmann, T., and Barandiaran, X. E. 2017. *Sensorimotor life: an enactive proposal*. Oxford University Press, Oxford.

[8] Mitchell, M. 2011. *Complexity: A Guided Tour*. Oxford University Press, Oxford.

[9] Jug, J., Kolenik, T., Ofner, A., and Farkaš, I. 2018. Computational model of enactive visuospatial mental imagery using saccadic perceptual actions. *Cognitive Systems Research*, 49 (2018), 157–177. DOI= https://doi.org/10.1016/j.cogsys.2018.01.005.

[10] Kolenik, T. 2016. Embodied Cognitive Robotics, the Question of Meaning and the Necessity of Non-Trivial Autonomy. In *Cognitive science: proceedings of the 19th International Multiconference Information Society – IS 2016* (Ljubljana, Slovenia, October 13, 2016). Jožef Stefan Institute, 24–27.

[11] Oyama, S. 1985. *The ontogeny of information: developmental systems and evolution*. Cambridge University Press, Cambridge.

[12] Kolenik, T. 2018. Seeking after the Glitter of Intelligence in the Base Metal of Computing: The Scope and Limits of Computational Models in Researching Cognitive Phenomena. *Interdisciplinary Description of Complex Systems*, 16, 4 (2018), 545–557.

The state of the Integrated Information Theory, its boundary cases and the question of 'Phi-conscious' AI

Tine Kolenik
Jožef Stefan Institute & Jožef Stefan International Postgraduate School
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3807
tine.kolenik@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3644
matjaz.gams@ijs.si

ABSTRACT
This work analyzes Giulio Tononi's Integrated Information Theory of consciousness, defined in 2016, the tools it offers to calculate the level of consciousness in any given system, produced in 2018, and compares the theory to other relevant recent theories of consciousness. It then discusses issues with the theory as well as the tools, namely that they are unreliable due to a variety of shortcuts that give different approximations, as current technology does not allow faithful computation of consciousness, i.e. a system's Phi. The testing confirms the problems with running time (*O*). Tononi's stand on AI is then problematized in relation to IIT. The authors' thoughts and treatise on a possibility of Phi-conscious AI is presented afterwards. AI systems are separated in three levels of hierarchy according to Marr and two types – knowledge representation-based and neural network systems according to Shoham. The authors hypothesize that combining both types brings AI closer to consciousness, which should hold true according to the multiple knowledge principle. Both systems are evaluated in relation to IIT's axioms and postulates. Evaluation shows that their combination conforms to more axioms and postulates than both types do separately, therefore confirming the hypothesis. However, AI is still not Phi-conscious as it does not encompass all of IIT's requirements.

Keywords
Artificial intelligence, consciousness, functionalism, Integrated Information Theory.

1. INTRODUCTION
Consciousness, this infinitely intimate state that we cannot escape and which encompasses our every thought, our every feeling and our every experience, is currently one of the most explored phenomena in science. It was explored with natural scientific methods more than 100 years ago by figures like the psychophysicists William James, Gustav Fechner, Hermann von Helmholtz and Wilhelm Wundt, but the research stopped as it was seen as a primitive, subjective and unscientific practice [1]. However, since the late 1990s, consciousness was again established as a phenomenon not only worth of exploring, but being able to be explored [2].

Theories of consciousness are abound, and there are many unique proposals, featuring orthogonal presuppositions, various ontological claims and sequestered methodologies for inquiry. Some of the most well received recent theories include the Global Workspace Theory [3], the Multiple Drafts Model [4], predictive coding approaches [5] and quantum theories of consciousness [6]. Among all, the Integrated Information Theory (IIT) of consciousness [7], proposed by the neuroscientist and psychiatrist Giulio Tononi, was described as the most formally sound, most computer science related and the most scientifically viable theory in this field yet [8].

IIT is based on a mathematical concept or quantity Φ, Phi, which can be calculated for any given system and represents *integrated information* (more in Section 3). IIT claims that integrated information is almost entirely correlated with the level of consciousness in the system Φ is calculated for. For example, the human brain has a very high Φ, which according to IIT, means that it is very highly conscious. But Φ can be calculated for any given system, so even atoms have some low number of Φ, or systems such as a light switch [9]. This conceptualization comes very close to the philosophical view of the mind called panpsychism, which proposes that consciousness or mind is a fundamental property of each and every part of any given system (from atoms to rocks to buildings to planets to the universe itself) [10]. This connection was also acknowledged by Tononi and Koch [11]. Another important aspect of IIT pertains to the hard problem of consciousness, which describes the explanatory gap between qualia or experience and physical states. IIT eschews the hard problem by presupposing consciousness as intrinsically real due to a system's cause-effect powers upon itself (see Section 3, Axiom 1). This axiomatic property of IIT circumvents the hard problem debate, which is why it will also not be addressed any further in this work as it is out of its scope. The wider framework of IIT is described in Section 3. However, since even photodiodes' Φ is above zero, the threshold for levels of semantically reasonable consciousness should be above zero in order to differentiate between what is commonly seen as conscious and unconscious. This should serve for easier discussions on consciousness in boundary cases such as artificial intelligence (AI).

In general, this paper is an upgrade of the paper by Gams [12], who presents an older version of IIT defined in 2014, offers a commentary on it and sets foundations for discussing AI in relation to IIT. The current work encompasses:

a. the state of the mentioned recent theories on consciousness in order to set them apart from IIT (Section 2),

b. an analysis of the state of IIT in its updated, newest form alongside with the recently developed tools

available for measuring consciousness of any given system (Section 3), and

c. an analysis of the boundary cases for consciousness as described by Tononi [13] with the focus on AI and its possibilities for possessing consciousness (Section 4).

The paper ends with the authors' intentions for future work and some concluding thoughts.

2. STATE OF THE RECENT THEORIES OF CONSCIOUSNESS

This Section briefly presents the current state of the following theories on consciousness: the Global Workspace Theory [3], the Multiple Drafts Model [4], predictive coding approaches [5] and quantum theories of consciousness [6]. It also offers a short criticism of each and whether they encompass the possibility for AI to be conscious.

The Global Workspace Theory (GWT), which spawned many advanced off-shot theories such as the 'neuronal global workspace' theory [14], relies on the concept of global availability of conscious content. Conscious content is supposedly available to all cognitive processes (e.g., attention, decision-making), which are connected more to certain parts of the brain, while conscious content inhabits a global neuronal activity across the brain. Consciousness is therefore widely spread, while various processes and states compete for being brought into this conscious landscape. The theory can explain various neuronal phenomena as well as functional cognitive processes, but it is not clear on how the graduality (or binariness) of consciousness works and how to precisely measure it. If the organizational aspects of GWT were realized in computers, it would be sensible to say that computers would be conscious.

The Multiple Drafts Model is a cognitivist theory of consciousness and proposes that there is "no reality of conscious experience independent of the effects of various vehicles of content on subsequent action (and hence, of course, on memory)." [4, p. 132] The theory claims that there are numerous interpretations of the sensory data that comes in through our senses. Since these are processed in different parts of our brains at different times, the first of the multiple drafts that checks all the necessary boxes in the neural processing is the one that is acted upon, and that the experience accompanying it is illusory. However, critics claim that the theory does not hold the power to explain or predict neuropsychological research data. It also does not offer mathematical explanations. Regardless, Dennett believes that mental functions are functions in a mathematical sense, which means that they can be formalized in a machine, resulting in a conscious AI.

Predictive coding approaches [5] are probably the most recent approaches to understanding the mind. Predictive coding refers to the theory that the minds and brains are fundamentally prediction machines. The mind builds a hierarchical generative model of the world which it is always predicting. This radically changes the idea that the sensory input and information-processing of it is a feed-forward process, that sensory data travels from, e.g., the eye through the brain's multiple layers of processing, and in the end, causes a motor action. Instead, the brain predicts the next input to the eyes before the input appears. The theory is currently one of the most researched, if not the most researched theory in cognitive science [15]. Predictive coding is a highly mathematical theory, as it partly relies on computer science algorithms, meaning that it should be able to encode at least some aspects of what predictive coding has to say on consciousness in machines.

Quantum theories of consciousness mainly claim that classical mechanics cannot explain consciousness. It is quantum entanglement and superposition as well as other quantum phenomena that cause consciousness [6]. However, the quantum hypotheses mostly discuss how quantum phenomena may give rise to consciousness and not much about the consciousness itself. The main (and particularly enormous) problem is that they are nowhere near testable. Since the quantum theories rely on quantum phenomena in terms of consciousness existing, machines first need to possess these quantum phenomena. Then, according to the theory, they can be built to have consciousness.

This collection of various contemporary theories of consciousness tries to sketch the state of consciousness theories so that IIT is placed in context and that it can be evaluated against them. The next Section discusses the state of IIT.

3. STATE OF THE INTEGRATED INFORMATION THEORY

This Section more thoroughly introduces IIT and the recently released tools and methods for measuring Φ. This serves as a continuation and an upgrade of the description of IIT by Gams [12] as well as a foundation on which Section 4 analyzes AI in regards to Φ.

The IIT takes inspiration from various sources – panpsychism was already mentioned – but it starts from getting away from purely searching for neuronal and behavioral correlates of consciousness and experience. It asks the harder questions of why cerebral cortex gives rise to consciousness but not cerebellum, even though it has approximately 4 times more neurons than the cerebral cortex and of what is important for consciousness in terms of various boundary cases having it. The latter is especially important, and Tononi and Koch [11] list a number of such cases where they ask whether they are conscious or not: 1) patients and infants, 2) animals, and 3) machines (more on this in Section 4). IIT therefore does not want to only work with collected data on cases where consciousness is freely attributed – neurotypical adult humans – it wants to propose what consciousness and experience are and what kind of systems in regards to their interactional properties can have them. IIT does that, however, in a reverse order than what consciousness researcher usually do – it starts from experience by positing five axioms and deriving five postulates that describe systems for which the axioms are true. On top of that, IIT establishes a calculus for precise measurements of consciousness, which it connects to integrated information, symbolized by Φ, Phi.

The five axioms and postulates are:

1. Intrinsic experience:

Axiom: Consciousness is real, and it is real from its own perspective.

Postulate: System must have cause-effect power upon itself.

2. Composition:

Axiom: Consciousness is composed of phenomenological distinctions, which exist within it.

Postulate: System must be composed of elements that have cause-effect power upon the system.

3. Information:

Axiom: Consciousness and each experience is specific, differing from other possible experiences.

Postulate: System must possess cause-effect sets that differ from each other in their space of possibilities.

4. Integration:

Axiom: Consciousness is unified and experience is irreducible to a set of its phenomenological distinctions taken apart.

Postulate: System must specify its cause-effect structure as to be unified, irreducible to mere sum of its parts ($\Phi_{system} > \Phi_{sum\ of\ parts}$).

5. Exclusion:

Axiom: Consciousness and experiences are definite and are the way they are, nothing else.

Postulate: System must specify its cause-effect structure to be definite, always over a single set of elements and maximally irreducible ($\Phi_{system} > \Phi_{any\ given\ sub\text{-}system}$).

The remaining part of this Section focuses on the notion of integrated information, Φ, as this is the part of IIT that Tononi's team is paying attention to the most in the recent years in terms of updating and revising it, especially with new tools.

Among others, the notion of integrated information offers the answer to the question of why cerebral cortex generates consciousness, but not cerebellum, even though the later has four times more neurons than the first. It also explains how even photodiodes can have experience and therefore, albeit very low level of, consciousness.

The main idea behind Φ and why it measures consciousness is this: First, it measures information in a certain system. This information is denoted by how much information the system has about itself, which is defined as a number of possible states, past and future. Second, this measure of information is coupled with how this information is integrated. What is measured is how much the information depends on the interconnectedness of the system's parts. To demonstrate this measurement, the system is split (into an arbitrary number of sub-systems) and then information is measured again. The more information that is lost, meaning the more information that arose from this interconnectedness, the more integrated the system was. Integration is also the reason why Tononi argues that computers have very little consciousness – because even though they can have much information, it is not integrated. He argues that transistors (he deems the physical, implementational level the most important) do not lose much structure or information if split, as they can still give rise to the same system (more on this in the next Section).

However, measuring Φ, even if we generally know what we want to measure, is extremely difficult. The biggest problem is that Φ cannot be calculated with our current computational technologies even if the system is only as big as a few nodes. Φ can be approximated with various different shortcuts and heuristics, but the problem is that for the same system, the approximation wildly varies depending on the technique for the approximation used [16]. In 2018, Mayner et al. [17] produced PyPhi, a Python software library that allows one to study the cause-effect structure of a given system in relation to IIT and calculate Φ. However, even though it encompasses a number of heuristics to calculating Φ, the algorithm's running time is exponential in terms of number of nodes increasing. Currently, the algorithm's running time is $O(n53^n)$, where n denotes the number of nodes. Running simple CPU experiments, it takes 24 hours to calculate Φ using the major complex of systems approach on a seven-node system if run on 4 × 3.1GHz CPU cores (see Table 1). Other shortcuts produce different running times, but also different Phis.

Table 1: Test of running time of Φ calculations for three systems with a different number of nodes.

# of nodes in system	Running time
3	~8 seconds
5	~2.5 minutes
7	~24 hours

The running time and the problem of getting different Phis with different calculations is one of the biggest criticisms of IIT. It also seems that in its current version, V3, IIT does not provide falsifiable predictions, which is one of the most common criticisms of most theories of consciousness.

4. INTEGRATED INFORMATION THEORY AND ARTIFICIAL INTELLIGENCE

This Section speculates on conscious AI in relation to IIT, dubbed as Phi-conscious AI. The authors address some of Tononi's points on AI, argue that some of his points may not be correct regarding it, propose that AI on certain levels may be seen as conscious and evaluate different AI paradigms through IIT's axioms.

Tononi examines AI only from a physical level. He only considers what computers are physically made of and makes claims exclusively about transistors and their inability to reach high Φ due to not being integrated – if one splits transistors, they can still possess the same information value. Tononi even states that if "integrated information theory is correct, computers could behave exactly like you and me, and yet there would literally be nobody there" [18, para. 32]. This means that even if they were programmed to satisfy the axioms and have a sufficiently high Φ, according to Tononi, their physical, transistor-based implementation would preclude 'true' consciousness. AI that would behave perfectly humanly would be the philosophical zombie. However, Tononi takes a very narrow perspective on AI that may even be in contention with IIT itself, as IIT's axioms and postulates do not necessarily require the implementational level of a system to be the one that counts in term of consciousness. Marr [19] proposes a three-level hierarchy in regards to AI and cognition in general: 1) computational level (what the system does and why), 2) algorithmic level (how the system does what it does), 3) physical level (the realization of the first two levels). The first two levels may bear a much higher Φ. However, the computational level does presuppose some functionalist ideas, namely that mental states are as they are because of the function they perform.

To speculate on whether certain types of AI on the 1st and 2nd level of Marr's hierarchy are Phi-conscious, AI is separated in three categories. It is investigated whether IIT's axioms and postulates hold true for them. The AI categorization is based on Yoav Shoham's invited talk [20] at this year's International Joint Conferences on Artificial Intelligence (IJCAI), one of the biggest

and oldest AI conferences in the world. Shoham categorizes AI in roughly two categories: knowledge representation (KR) based AI (commonly dubbed as 'good old-fashioned AI') and neural networks (NN). His hypothesis is that KR is good for certain problems, that NN is good for other problems and that by combining the two, AI will enter a new era of progress as KR+NN will work better than its parts (see Figure 1). Our hypothesis mirrors Shoham's – we believe that KR may satisfy some IIT's axioms and postulates, that NN may satisfy some other axioms and postulates, but that together they would have higher Φ than they would if treated separately and then summed up. This thinking is also based on the multiple knowledge principle [21], according to which our hypothesis should hold true.

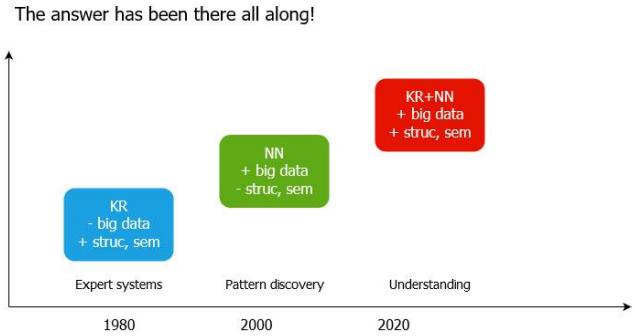


Figure 1: Shoham's vision for AI (struc = structured, sem = semantics). Adapted from [20].

KR mostly encompasses expert systems. These are systems that have all their domain knowledge programmed into them with various rules, which are explainable and symbolic in nature. The process of knowledge acquisition is top-down, meaning that the designer presupposes everything they know.

NN encompass learning systems that usually look for patterns. Their knowledge is produced from lots of data (big data), bottom-up, they are subsymbolic and very robust.

The table below (Table 2) shows the analysis for KR, NN and KR+NN in relation to IIT's axioms and postulates. KR+NN's relation to IIT is determined by using logical disjunction (∨, (x)or) between KR and NN, as axioms and postulates have to hold true only for one to hold true for KR+NN.

The arguments in Table 2 claim that by combining KR and NN, AI gets closer to achieving consciousness according to IIT. What seems to be lacking in both is *exclusion*. AI therefore cannot be characterized as being Phi-conscious just yet, but our initial hypothesis is confirmed.

There is more to IIT's problems regarding AI. One problem is that Tononi clearly states that his theory should be judged according to how it explains the empirical data about consciousness [11]. There is a problem with this in relation to AI – there is no empirical data about consciousness. Tononi presupposes consciousness and acts accordingly – that neurological data on the brain is in fact empirical data about consciousness, without calculating Φ to find out whether this is true. This inherently cripples meaningful research on AI consciousness, as one cannot do the same and presuppose it in, e.g., robots. You cannot, as Tononi tries to do with IIT, reverse engineer the process of scientific investigation and theorizing.

Table 2: Analysis of KR, NN and KR+NN in relation to IIT's axioms and postulates.

AI type / IIT	KR	NN	KR+NN (KR ∨ NN)
Intrinsic experience	can have cause-effect power upon itself, as rule-based system may operate on feedback loops and recursions (the specified rules may change) that are being performed without input **TRUE**	layers may easily be interconnected or connect in a way (bi-directional layers, feedback loops on the same nodes …) for NN to have cause-effect power upon itself, especially in no-input NNs such as (generative NN, Kohonen NN …) **TRUE**	**TRUE**
Composition	has strong compositional property; computational rules may be linked between each other and have effect among each other **TRUE**	due to the self-organizational nature of NNs, modularity and therefore composition is not clear and entirely explainable; nodes do connect, but may not hold true for concepts; since it is very robust, parts may be removed without affecting the system itself **FALSE**	**TRUE**
Information	can possess many cause-effect sets, differing from each other (Tononi also states that machines have high information value) **TRUE**	a number of cause-effect sets is usually operationally the same in relation to their power in the system (which is why optimization by reducing NN size works) **FALSE**	**TRUE**
Integration	in KR, the sum of its parts by definitions cannot be more than the system itself, as expert systems are inherently modular, therefore violating '$\Phi_{system} > \Phi_{sum\ of\ parts}$' **FALSE**	works as a unified and distributed system and completely irreducible to the sum of its parts as nodes necessarily organize between each other in an inseparable way; '$\Phi_{system} > \Phi_{sum\ of\ parts}$' holds true **TRUE**	**TRUE**
Exclusion	cannot guarantee that a KR system is a maximally irreducible, especially due to its modularity, therefore violating '$\Phi_{system} > \Phi_{any\ given\ sub\text{-}system}$' **FALSE**	Usually a NN can be reduced to an operationally equally effective subsystem that has the same integration and information values (which is why optimization by reducing NN size works), which implies that NN systems violate '$\Phi_{system} > \Phi_{any\ given\ sub\text{-}system}$' **FALSE**	**FALSE**

5. CONCLUSIONS AND FUTURE WORK

This work presents the latest iteration of the Integrated Information theory proposed by Tononi, some tools the IIT researchers offer for calculating Φ, and the problems of both. Some other theories of consciousness are presented as well to put IIT in context, especially in regards to AI. The biggest contribution of this work is in trying to speculate on whether AI is, as dubbed by the authors, Phi-conscious or not. We speculate about consciousness on various types of AI, categorized by Shoham, and hypothesize that combining different types brings us closer to Phi-conscious AI, which we claim to confirm (Table 2).

Our future work includes more thorough analysis of different concrete KR and NN systems, but our foremost interest lies in working with KR+NN systems. This seems to be the future regardless of IIT, but we want to make KR-NN systems as close to Phi-conscious as possible and see what consequences will emerge. Other ideas for future work include using machine learning and state-of-the-art algorithms to deal with the algorithm running time better in terms of developing heuristics to shorten the calculating time, and consequently calculating Phi for systems such as recurrently connected Turing machines to find out whether it is higher than the sum of individual Turing machines due to dynamic interactions [21].

6. REFERENCES

[1] Hawkins, S. L. 2011. William James, Gustav Fechner, and Early Psychophysics. *Front. Physiol*. 2, 68 (2011). DOI= 10.3389/fphys.2011.00068.

[2] Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press, Oxford.

[3] Robinson, R. 2009. Exploring the "Global Workspace" of Consciousness. *PLoS Biol*. 7, 3 (2009), e1000066. DOI= 10.1371/journal.pbio.1000066.

[4] Dennett, D. C. 1991. *Consciousness Explained*. Little, Brown & Co, Boston, MA.

[5] Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci*. 36 (2013), 181-204. DOI= 10.1017/S0140525X12000477.

[6] Atmanspacher, H. 2019. Quantum Approaches to Consciousness. *The Stanford Encyclopedia of Philosophy*, (2019).

[7] Tononi, G., Boly, M., Massimini, M., and Koch, C. 2016. Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci*. 17, 7 (2016), 450-461. DOI= 10.1038/nrn.2016.44.

[8] Gennaro, R. J. 2018. *Consciousness*. Springer, Cham.

[9] Tononi, G. 2008. Consciousness as Integrated Information: a Provisional Manifesto. *The Biological Bulletin*. 215, 3 (2008), 216-242. DOI= 10.2307/25470707.

[10] Alter, T. and Nagasawa, Y. 2015. *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press, Oxford.

[11] Tononi, G. and Koch, C. 2015. Consciousness: here, there and everywhere? *Phil. Trans. R. Soc. B*. 370 (2015). DOI= 10.1098/rstb.2014.0167.

[12] Gams, M. 2015. Kochove meritve zavesti. In *Cognitive science: proceedings of the 18th International Multiconference Information Society – IS 2015* (Ljubljana, Slovenia, October 8-9, 2015). Jožef Stefan Institute, 11-14.

[13] Oizumi, M., Albantakis, L, and Tononi, G. 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol*. 10, 5 (2014), e1003588. DOI= 10.1371/journal.pcbi.1003588.

[14] Dehaene, S. 2015. *Consciousness and the Brain*. Viking, New York.

[15] Bohannon, J. 2018. *A computer program just ranked the most influential brain scientists of the modern era*. Retrieved September 11, 2019, from https://www.sciencemag.org/news/2016/11/computer-program-just-ranked-most-influential-brain-scientists-modern-era.

[16] Bayne, T. 2018. On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 2018, 1 (2018), niy007. DOI= 10.1093/nc/niy007.

[17] Mayner, W. G. P., Marshall, W., Albantakis, L, Findlay, G., Marchman, R., and Tononi, G. 2018. PyPhi: A toolbox for integrated information theory. *PLoS Comput. Biol*. 14, 7 (2018), e1006343. DOI= 10.1371/journal.pcbi.1006343.

[18] Robson, D. 2019. Are we close to solving the puzzle of consciousness? Retrieved September 11, 2019, from http://www.bbc.com/future/story/20190326-are-we-close-to-solving-the-puzzle-of-consciousness.

[19] Marr, D. 2010. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press, Cambridge, MA.

[20] Shoham, Y. *Award for Research Excellence presentation. Invited Lecture*. 2019 International Joint Conference on Artificial Intelligence.

[21] Gams, M. 2001. *Weak intelligence: through the principle and paradox of multiple knowledge*. Nova Science, Hauppauge, NY.

Establishing illusionism

Alen Lipuš
University of Maribor
Trg revolucije 7
2000 Maribor
+38641415441
alen.lipus1@um.si

Janez Bregant
University of Maribor
+38641658397
janez.bregant@um.si

ABSTRACT
In his recent paper on the meta-problem of consciousness, Chalmers claims that illusionism is one of the best reductionist theories available and that it is not incoherent even if it is implausible and empirically false. The paper argues against this. The first part introduces the reasoning leading to illusionism, i.e. it describes the initial conditions and relations among them for establishing it. The second part of the paper argues that strong illusionism is not set up in a satisfactory way and calls the flaw in establishing it the pre-illusion problem.

Keywords
Consciousness, illusionism, phenomenal properties, knowledge argument, first-person perspective.

1. INTRODUCTION
When we reflect on what it means to be conscious or what it means to undergo a certain qualitative experience we are faced with the following problem: the subjective aspect of the first-person experience is not compatible with physicalism. Traditionally, phenomenality is understood as a cluster of 'what it's like' properties that determine the phenomenal character of a mental state. There is a consensus among most philosophers that the phenomenal states threaten the truth of physicalism. The phenomenal cluster consists of phenomenal properties being, among other things, ineffable, irreducible, intrinsic, direct, subjective, private etc. So, the problem of relating such properties to something purely physical emerges naturally: How does conscious experience emerge from physical processes in the brain? The problem framed this way and called by Chalmers the hard problem poses a great threat to any physicalist strategy [1]. In the contemporary philosophy of mind, the discussion regarding the hard problem has been radicalized to the point that mainstreem traditional physicalism is losing its proponents. We see fewer philosophers who are ready to maintain a compatibilist position, that mental, phenomenal states are real and can be placed within the physicalist ontology. On one hand we have the realists about mental states, who maintain that the placement problem of mental states is indicative of their special nature, namenly thier non-physical nature [1, 2, 7, 11]. Since we cannot fathom how can mental states, if real, be placed within the physical framework, this means that the mental states must somehow be something extra-physical. On the other hand we have philosophers, who realized that one cannot be a realist about mental states and at the same time hold that physicalism is true, and therefore their physicalist position is radicalised to the point that they deny the reality of mental states [4, 5, 9, 10].

One of these strategies is called illusionism. It does not try to solve the hard problem but to dissolve it by showing that something like phenomenality as described does not exist at all. And, if there is no phenomenality then there is no hard problem of consciousness. Chalmers sees it as the best reductionist approach to the explanation of consciousness [2]. According to his line of thinking, what we are left with is the so-called illusion problem: "Why does it seem that we have phenomenality when we really don't" [5]. There are several answers to the question of how the illusion of phenomenality[1] arises but they will be left aside [8, 9, 14]. The focus of the paper is on the reasoning leading to illusionism, more precisely, the evaluation of the contemplation process establishing the illusionist position. Firstly, the paper describes how strong illusionism is set up. Secondly, it argues that there is a flaw in setting it up called the pre-illusion problem.

2. SETTING-UP ILLUSIONISM
We introduce the illusionistic modifications to phenomenality as uncovered by Frankish through a simulation of the reasoning leading to illusionism [5]:

1. Phenomenality is/seems anomalous.

2. A commitment to an explanatory strategy that relies on existing theoretical resources without major revisions.[2]

∴ (3) Phenomenality does not exist.

The first premise is understood as "phenomenality is anomalous" by strong illusionism and as "phenomenality seems anomalous" by weak illusionism. Weak illusionism claims that the mere possibility that phenomenality is anomalous is already enough and ties it to certain suspected anomalous characteristics that phenomenal states possess, i.e. they are private, ineffable, immediately apprehended, intrinsic, direct. However, some authors think that strong illusionists are right in saying that weak illusionism either collapses into strong illusionism or it cannot do the job that it sets out to do [5]. The second premise emphasizes the importance of relying on existing theoretical resources, its mantra is "first exhaust, then propose" [4, 5]. According to this methodology, one should deal with a problem by, firstly, trying to exhaust all the existing theoretical resources, and, secondly, making radical theoretical revisions (the second step is made only in case of the failure of the first one). The exhaust/propose approach is somewhat straightforward as it is present even in the radical realist camp.[3]

[1] There are no instantiated phenomenal properties.

[2] Major theoretical revision would be some metaphysical modification of physicalism to accommodate phenomenal consciousness, e.g. panpsychism, where consciousness is a fundamental property of matter.

[3] Those who are already making radical theoretical revisions and are modifying the existing metaphysics in a nonphysical way

Nevertheless, according to strong illusionism, the fact that phenomenality (as standardly characterized) is anomalous and since physicalists (realists about phenomenal states)[4] have a problem explaining phenomenal consciousness, illusionism lends itself as a good radical explanation of phenomenality. To preserve physicalism, it must explain phenomenal states as illusory [3, 5].[5] One of the best course of action in dealing with anomalous phenomena is to declare them illusions, especially if one has good reasons to stay committed to the current explanatory framework provided by physical sciences. This way illusionists do not banish consciousness but modify it to fit the physicalist world. On their view, conscious states do not possess real phenomenality but merely, the so-called, quasi-phenomenality [5]. These quasi-phenomenal properties are functional properties of brain states. We get tricked by consciousness[6] as our introspective self-representation mischaracterizes the physical/functional properties as phenomenal. There really are no phenomenal properties instantiated in our mental states, we only wrongly think that the essential characteristic of consciousness is 'what it's like'. The research project for illusionism is, therefore, to explain and identify mechanisms that are responsible for phenomenal misattribution.

As far as the hard problem is concerned, its position is obvious and very straightforward: there is no such problem because there is no true phenomenality.[7] The next step is to explain why then we are prone to phenomenal judgements,[8] why we think that we are phenomenally conscious, and why the illusion of phenomenality is so powerful. There are already several theories that deal with the questions at hand: some identify the underlying firmware of our introspection as a candidate for the misattribution [8, 9, 14], some find the perpetrator in the flawed inferential mechanism [10], and some combine the misaligned introspective mechanism with philosophical prejudices [4] in order to account for the misattribution. Still, what we are concerned with in this paper is not an answer to the question of why the illusion of phenomenality arises but with identifying a mistake in the sheer concept of illusionism. Because the incoherence in conception can be a source for the incoherence in perception, what is called the meta-illusion problem [13], we will analyze the initial establishing conditions of illusionism.

3. INCOHERENCE OF ILLUSIONISM

Illusionism sees phenomenality in general to be incompatible with physicalism and, therefore, turns it into quasi-phenomenality that is supposed to align with physicalism. In what follows, we are not going to argue for such functional transformation of phenomenal properties but are going to show that illusionism is built on false initial assumptions. We will introduce the central thesis (T) of our argument first and then work backwards to construct it.

T*: To be justified in denying phenomenality, one must accept the claim that phenomenality exists.*

It is a puzzling situation for illusionism as the following question nicely shows: If there really are no such things as phenomenal states how do we know that they are incompatible with physicalist metaphysics? One of the essential characteristics of phenomenal consciousness is that we must have the first-person perspective 'what it's like' experiences to know that they have a phenomenal character. There is no other way to know what something is phenomenally like but to have a private subjective experience of it. And this is exactly the feature of phenomenality that threatens to reject physicalism once and for all. The famous Knowledge argument [11, 12] is one strong example of how to dismiss physicalism on the 'what it's like' ground. Phenomenal states have a devastating characteristic from the physicalist/illusionist point of view: they are by their nature the first-person perspective states. No amount of careful speculation and imagination can reveal what they are like. This characteristic is what makes them anomalous and it is what gives such a striking power to the hard problem of consciousness. We get to know what phenomenal states are by having 'what it's like' subjective experience of them, and illusionists are no exception. Yet, someone might say that our objection does not affect illusionism since they deny the existence of the phenomenal character of experiences, i.e. there is no 'quale' involved in no matter what mental states. It is clear why illusionists have to refuse it, but the question is how can they dismiss something, i.e. 'what it's like', without experiencing it? Given the nature of phenomenal states, they cannot. And does not then having the subjective qualitative experience mean that something like phenomenality must exist before it is denied? Given the nature of phenomenal states, it must. We call this the pre-illusion problem. Let us now recapitulate the story of how someone becomes an illusionist. First, she has something like phenomenal experience whose nature is, in the light of physicalism, anomalous, which generates the hard problem. Second, since she wants to keep the theoretical advantages of the physicalist explanatory repertoire, the only natural thing to do seems to reject the existence of phenomenality and to become the illusionist. But to deny phenomenality illusionists must have the first-person perspective experience of it, they must be subjectively acquainted with it. How else would they know that phenomenality is anomalous? Illusionists cannot say that phenomenal states are not revealed through phenomenal experience, or that they are not tied to the first-person perspective experience since the elimination of their supposed properties undermines the case for strong illusionism: if phenomenal states do not have these characteristics then they are not anomalous and the motivation for illusionism is lost. But what is in the first place that is anomalous? It seems that to conceptualize the anomalous nature of phenomenal experience one must first have it: we cannot conceptualize the phenomenal character of mental states in any other way, and this is exactly what makes phenomenality anomalous. Moreover, why would physicalists deny the existence of phenomenality if it did not have the problematic 'what it's like properties' that makes it anomalous? It turns out that

simply follow the described methodology: physicalism is exhausted so bring out some new, i.e. nonphysical, explanation of phenomenality.

[4] E.g. phenomenal concepts strategy

[5] The analogy drawn here is the one with paranormal powers, such as telekinesis. The phenomenon of telekinesis is anomalous to our scientific understanding of the world; thus, we can modify the naturalistic framework to accommodate telekinesis or we explain it away as an illusion.

[6] Consciousness can be understood in functionalist terms, e.g. access consciousness, where a mental state is not qualitatively present to the organism, but it is generally available to it**.**

[7] In other words, phenomenal consciousness does not need to be explained since it does not exist, i.e. there is no phenomenal consciousness instantiated in our world. This is the so-called meta-approach (denying or questioning the hard problem) to the explanation of consciousness within the physicalist framework.

[8] Chalmers calls them phenomenal reports [2].

strong illusionism is left with the catch-22 situation:[9] on the one hand it refuses the existence of phenomenal states, but on the other hand it accepts it to be justified in denying them. However, we are not justified to reject something that exists, therefore strong illusionism, as it is set up now, is not a well-founded theory.

4. CONCLUSION

We introduced the pre-illusion problem as a real threat to the truth of illusionism because it prevents it from being established in the first place. It shows that the argumentation leading to a creation of illusionism is flawed: to know that phenomenal properties are anomalous requires to be subjectively familiar with them, i.e. to experience their 'what it's like' from the first-person perspective, a condition that is not met by illusionism. The very anomalous nature of phenomenal properties, the one that is incompatible with physicalism, is not a reflective by-product of our metaphysical imagination but something that we experience. Illusionism can be seen as a good dialectical position; it recognizes the metaphysical allure of phenomenality and tries to save physicalism by turning the phenomenal nature of mental states into the functional one. Unfortunately, it seems that to get to know the anomalous nature of phenomenal properties we must undergo qualitative private experiences, which renders a denial of phenomenality by illusionists impossible. This means trouble, so the pre-illusion problem must be solved if they want their theory to be plausibly established at all.

5. REFERENCES

[1] David J. Chalmers. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press

[2] David J. Chalmers. 2018. The Meta-Problem of Consciousness. (January 2018). Retrieved September 18, 2019 from https://philpapers.org/archive/CHATMO-32.pdf

[3] Daniel C. Dennett. 1991. *Consciousness Explained*. New York: Little, Brown.

[4] Daniel C. Dennett. 2005. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

[5] Keith Frankish. 2016. Illusionism as a Theory of Consciousness. (January 2016). Retrieved September 18, 2019 from https://www.ingentaconnect.com/contentone/imp/jcs/2016/00000023/f0020011/art00002

[6] Jay L. Garfield. 2015. *Engaging Buddhism: Why it Matters to Philosophy*. Oxford: Oxford University Press.

[7] Phillip Goff. 2017. *Consciousness and Fundamental Reality*. Oxford: Oxford University Press

[8] Michael Graziano. 2013. *Consciousness and the Social Brain*. New York: Oxford University Press.

[9] Nicholas Humphrey. 2011. *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.

[10] Georges Rey. 2016. Taking Consciousness Seriously – as an Illusion. (January 2016). Retrieved September 18. 2019 from https://www.ingentaconnect.com/contentone/imp/jcs/2016/00000023/f0020011/art00016

[11] Frank Jackson. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly.* 32, 127 (Apr. 1982), 127-36. DOI:10.2307/2960077.

[12] Frank Jackson. 1986. What Mary Didn't Know. *Journal of Philosophy*. 83, 5 (May 1986), 291–295. DOI: 10.2307/2026143

[13] François Kammerer. 2017. Can You Believe It? Illusionism and the Illusion Meta-Problem. (January 2008). *Philosophical Psychology*. 31, 1 (Oct. 2017): 44–67. DOI:10.1080/09515089.2017.1388361.

[14] Derk Pereboom. 2011. *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.

[9] Catch-22 is a situation from which an individual cannot escape because of contradictory rules.

Artificial intelligence and pain: a promising future

Duska Meh
Faculty of Medicine
Linhartova 51, 1000 Ljubljana
+386 40 201 615
meh.duska@gmail.com

Dejan Georgiev
Faculty of Arts
Askerceva 2, 1000 Ljubljana
+386 40 743 226
dejan.georgiev@gmail.com

Metod Meh
IBM d.o.o.
Ameriska 8, 1000 Ljubljana
+386 40 456 885
metod.meh@gmail.com

ABSTRACT
Artificial intelligence and cognitive computing give hope or even promise that humans with augmented abilities will empower many unanswered questions and provide unprecedented opportunities in the quest for pain management. With improved connectivity, depth and breadth of comprehensive information, phenomena are easier to understand, and it is easier to implement an improved and ethically acceptable healthcare of suffering people. This ubiquitous and everyday phenomenon of artificial intelligence could be incapacitating, however, we believe the world would be happier and more creative with this exceptionally important though at present still unmanageable friend and co-worker.

Keywords
Artificial intelligence, Cognitive computing, Ethics, Medicine, Pain, Psychology.

1. INTRODUCTION
Computers continue to empower our life and everything important influencing our world. This is reasonable as far as human beings and technology cooperate in the transformation of global understanding of changeable momentary conditions. The era of cognitive systems that are helpful in deepening human knowledge and expertise is one of the greatest opportunities humankind ever thought or dreamed possible.

The digitalised future is forcing blinkered individuals to continuously complete their idealised "selves". Consciousness is partly based on the person's processing of information from external and internal worlds transferred to the mind (internal mental life) [1]. Individuals tend toward seeking and maintaining balance (homeostasis) within their internal environment, even when faced with external changes [2; 3]. Complex sensory discriminative, affective, evaluative and cognitive processes must detect deviations of values, which need accurate regulation. The most important information for individuals is connected with their health. The human mind is able to process limited amount of information and this amount does not essentially change with time [4]. The growing amount of subjective information, including subjectively most essential healthcare-related information, overflows any person's cognitive abilities. It is estimated that it would take a qualified person 150 hours each week to read every piece of content published in their field of interest. With the help of cognitive computing, this enormous and continuously growing pool of information could potentially be mastered [5].

In reality, contemporary circumstances overtake most insufficiently informed and too conservative people: artificial intelligence is seen as taking advantage of machine-learning techniques, such as artificial neural networks and its applications for diagnostics and healthcare management decisions [6]. Academic medical research has the opportunity to implement machine learning in health care. So far, health system information still seems manageable [7]. In the future, as predicted for 2020, the footprint of professionally collected data will double every 73 days [5]. Complete health information should be accepted and perceived as the most important for modern informed individuals. They should receive and understand information, i.e. cognitively process the incoming stimuli [8-10].

2. WHERE ARE WE?
The contemporary reality reflects an unfortunate and troubling trend: the aging population is, as expected, characterised by a growing number of individuals reporting highly subjective experiences that burden them [11; 12]. Slovenian endurance and patience have become proverbial, but chronic diseases – due to their multiorgan involvement and long-lasting progression – are repeatedly and prevalently incapacitating, which renders them subject to frequent complaints. The perceived inconsistency is associated with comprehensive peculiarity of pain [13-15].

3. PAIN
Current understanding of pain as a comprehensive multidimensional phenomenon goes beyond its important and generally accepted Merskey's, Melzack's and Wall's definitions [16-18]. Pain itself is a body's protective mechanism; a response could be partly defensive, but harmful stimuli are potentially dangerous and could seriously affects patients' normal lives [7]. This most ubiquitous somatosensation, multidimensional and multifunctional phenomenon, is undervalued, in spite of its importance and prevalence [11]. The unpleasant and burdened comprehensive process limits the functional status of pain-afflicted subjects and adversely influences their quality of life. Pain is also costly to society and increases healthcare costs. It has been discussed but real interventional plans have never been addressed. In general, societies, governments and funding agencies are insufficiently interested in providing money for enough research, teachers and professionals. If they were, they could be informed on the one hand by the published information and on the other hand by statistics. The management of pain requires more health care resources than the treatment of diabetes, heart diseases and cancer combined [19]. The comparison of health care costs of people who report pain, and those who do not report pain, discloses important distinctions between the two groups in terms of controlling health needs, demographic characteristics and socioeconomic status [20].

4. WHAT ABOUT PAIN?
For now, it is impossible to get away with the quest of pain. Helping people to live better with pain may be achievable. This conviction motivates "real" pain professionals that know and understand this multifaceted and unpleasant condition. Experience is either direct (own) or indirect (emphatic). The leading

inspiration and *condition sine qua non* (an indispensable condition) is the belief that pain is manageable. The accomplished fact has to be emphasised: diminution of the impact of pain stays, falls or persists on the enthusiasm, eagerness for knowledge and immense motivation of exceptional individuals and their exceptional co-workers.

5. PAIN MEDICINE AND ARTIFICIAL INTELLIGENCE

The incorporation of artificial intelligence and machine learning into the field of pain medicine is, from the clinical point of view, entirely a matter of the future, but at the same time a real-time availability. Clinical decision is one of the cornerstones of pain-puzzle and a computerised support system with cognitive computing could potentially be very useful in objectively impacting the field of health care.

The era of cognitive computerised health care, especially pain care, will bring together individualised professional research and transdisciplinary data from a diverse range of healthcare sources to redefine a path to personalized, transparent, integrated, transdisciplinary and high-quality care [21-23]. Artificial intelligence uses different data, classical unstructured and recent structured data. They fall into two major categories, i.e. machine learning techniques and natural language processing methods [24-26]. Machine learning techniques analyse structured data such as imaging, genetic and electrophysiological data. Natural language processing methods extract information from clinical notes, medical journals and books, proceedings, etc. and turn them into analysable and machine-readable structured data [27, 28]. Artificial intelligence techniques efficiently assist motivated pain professionals with raised awareness, who are motivated by clinical problems, their prediction and recognition, management, outcome prediction and prognosis evaluation.

Despite the increasingly rich artificial intelligence and the literature on healthcare, the published research mainly concentrates around a few disease types: diabetes, cancer, some nervous system diseases, cardiovascular disease and rarely pain [29-36]. In the foreseeable future, personalized healthcare and advanced personalized medicine will focus on the diagnosis, prognosis, and treatment of individuals. More sophisticated diagnostic and therapeutic health devices will be used to gather data and successfully manage the involved subjects.

6. WHERE ARE WE GOING?

In the foreseeable future sophisticated algorithms will “learn” features from a large volume of data and then use the obtained insights to assist clinical practice. The disciplines concerned with pain as unsolved problem are medicine, psychology, information and bio-technical disciplines. Some researchers hope that patterns of somatic activity might one day serve as a “link” to pain response, although exclusively biological data are far from being a comprehensive solution and comprehensive management of such a disturbed and changeable multidimensional phenomenon.

7. OUR INTENTION

Our great and extraordinary opportunity is the contribution of “pain psychologists” in the development of successful systemic human–computer interaction. We are building a network of intra–, inter–, cross–, multi– and transdisciplinary interactions with the help of contemporary technology that gives us the freedom to go everywhere and be at the same time part of the research group that is based on human–human interactions.

The goal of our transdisciplinary and “multidimensional” professional research group is to provide a set of data we have access to (our patients), and to indicate our decision-making path.

Aspects that should be inscribed as data are psychological, biological, sociological, etc. The currently available data on pain are anamnesis/history, psychological, algological and neurological examination, functional examination (e.g. psychophysical, electrophysiological, morphological examination), immunological examination, appropriate psychological tests, questionnaires and a battery of tests. These data are unstructured and structured.

Additionally, we will define a minimal set of data needed to select a successful diagnostic tool and explain the path to obtain diagnosis from the available data.

This should be the basis for next steps needed.

We have to get access to and the ability to handle ever larger data sets (bigger data sources) and an artificial intelligence system capable to run on our dataset. Then, with access to bigger data sets, we can start implementing an artificial intelligence system to become a valuable help to pain patients, their families, social networks and societies.

8. FINAL REMARKS

Last but not least (for us even most important): ethical issues [37-41]. The digital revolution is needed to address the broader ethical and societal concerns of new technologies. These high-priority areas need specific ethical guidance. Computerised health system is developing and subsequently, changing healthcare and people living inside or at the border zone of health and life.

We are absolutely convinced that living and working for people who suffer is worth it.

9. REFERENCES

[1] Dijkerman, H.C. and De Haan, E.H.F., 2007. Somatosensory processing subserving perception and action: Dissociations, interactions, and integration. *Behavioral and Brain Sciences 30*, 224-230.

[2] Schulkin, J., 2003. Rethinking homeostasis: Allostatic regulation in physiology and pathophysiology.

[3] Schulkin, J., 2004. Allostasis, homeostasis, and the costs of physiological adaptation Cambridge University Press, Cambridge.

[4] Keidel, W.D., 1979. Informationsverarbeitung. In *Kurzgefaßtes Lehrbuch der Physiologie*, W.D. KEIDEL and H. BARTELS Eds. Georg Thieme, Stuttgart, 1-13.

[5] Ahmed, M.N., Toor, A.S., O'neil, K., and Friedland, D., 2017. Cognitive computing and the future of health care: the cognitive power of IBM Watson has the potential to transform global personalized medicine. *IEEE pulse 8*, 3, 4-9.

[6] Shahid, N., Rappon, T., and Berta, W., 2019. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS One 14*, 2, e0212356.

[7] Li, K., Wang, J., and Wang, N., 2019. Research on intelligent management of pain. In *IOP Conference Series: Materials Science and Engineering* IOP Publishing, 042033.

[8] Duch, W., 2001. Facing the hard question. *Behavioral and Brain Sciences 24*, 187-188.
[9] Gray, J.A., 1995. The contents of consciousness: A neuropsychological conjecture. *Behavioral and Brain Sciences 18*, 4, 659-676.
[10] Zeman, A., 2001. Consciousness. *Brain 124*, 1263-1289.
[11] Elzahaf, R.A., Tashani, O.A., Unsworth, B.A., and Johnson, M.I., 2012. The prevalence of chronic pain with an analysis of countries with a Human Development Index less than 0.9: a systematic review without meta-analysis. *Current medical research and opinion 28*, 7, 1221-1229.
[12] Merskey, H. and Bogduk, N., 2012. Classification of Chronic Pain. Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms. In *IASP Press* IASP Press, Seattle.
[13] Finlay, B.L. and Syal, S., 2014. The pain of altruism. *Trends in cognitive sciences 18*, 12, 615-617.
[14] Steinkopf, L., 2017. The social situation of sickness: An evolutionary perspective on therapeutic encounters. *Evolutionary Psychological Science 3*, 3, 270-286.
[15] Steinkopf, L., 2016. An evolutionary perspective on pain communication. *Evolutionary Psychology 14*, 2, 1474704916653964.
[16] Melzack, R. and Wall, P.D., 1965. Pain mechanisms: a new theory. *Science 150*, 3699, 971-979.
[17] Merskey, H., 1968. Psychological aspects of pain. *Postgraduate Medical Journal 44*, 510, 297.
[18] Williams, A.C.D.C. and Craig, K.D., 2016. Updating the definition of pain. *Pain 157*, 11, 2420-2423.
[19] Henschke, N., Kamper, S.J., and Maher, C.G., 2015. The epidemiology and economic consequences of pain. *Mayo Clinic proceedings 90*, 139-147.
[20] Gaskin, D.J. and Richard, P., 2012. The economic costs of pain in the United States. *The journal of pain : official journal of the American Pain Society 13*, 715-724.
[21] Bandura, A., 2004. Health promotion by social cognitive means. *Health education & behavior 31*, 2, 143-164.
[22] Meh, D., Meh, K., and Georgiev, D., 2015. Medical Approach to Pain as Transdisciplinary Phenomenon. In *Proceedings of the Kognitivna znanost* (Ljubljana2015), Institut Jožef Stefan, 27-32.
[23] Georgiev, D., Meh, K., and Meh, D., 2015. Psychological approach to pain as transdisciplinary phenomenon. In *Proceedings of the Kognitivna znanost* (Ljubljana2015), Institut Jožef Stefan, 15-19.
[24] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., and Wang, Y., 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology 2*, 4, 230-243.
[25] Singer, M., Deutschman, C., Seymour, C., Wildman, M., Sanderson, C., Groves, J., Wildman, M., Sanderson, C., and Groves, J., 2019. Artificial intelligence in healthcare: past, present and future. *Journal of the Intensive Care Society 20*, 3, 268-273.
[26] Darcy, A.M., Louie, A.K., and Roberts, L.W., 2016. Machine learning and the profession of medicine. *Jama 315*, 6, 551-552.
[27] Elkin, P.L., Froehling, D.A., Wahner-Roedler, D.L., Brown, S.H., and Bailey, K.R., 2012. Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes. *Annals of Internal Medicine 156*, 11-18.
[28] Jha, A.K., 2011. The promise of electronic records: around the corner or down the road? *Jama 306*, 8, 880-881.
[29] Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine 25*, 1, 44.
[30] Courtine, G. and Sofroniew, M.V., 2019. Spinal cord repair: advances in biology and technology. *Nature medicine 25*, 898–908.
[31] Yuk, H., Lu, B., and Zhao, X., 2019. Hydrogel bioelectronics. *Chemical Society Reviews 48*, 6, 1642-1667.
[32] Schork, N.J., 2019. Artificial Intelligence and Personalized Medicine. In *Precision Medicine in Cancer Therapy*, D.D.V. HOFF and H. HAN Eds. Springer, 265-283.
[33] Sendak, M., Gao, M., Nichols, M., Lin, A., and Balu, S., 2019. Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities. *eGEMs 7*, 1, 1-4.
[34] Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., and Rhee, K., 2019. Transforming diabetes care through artificial intelligence: the future is here. *Population health management 22*, 3, 229-242.
[35] Collins, G.S. and Moons, K.G., 2019. Reporting of artificial intelligence prediction models. *The lancet 393*, 10181, 1577-1579.
[36] Hu, X.-S., Nascimento, T.D., Bender, M.C., Hall, T., Petty, S., O'malley, S., Ellwood, R.P., Kaciroti, N., Maslowski, E., and Dasilva, A.F., 2019. Feasibility of a Real-Time Clinical Augmented Reality and Artificial Intelligence Framework for Pain Detection and Localization From the Brain. *Journal of medical Internet research 21*, 6, e13594.
[37] Fiske, A., Henningsen, P., and Buyx, A., 2019. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of medical Internet research 21*, 5, e13216.
[38] Wilk, R. and Johnson, M.J., 2014. Usability feedback of patients and therapists on a conceptual mobile service robot for inpatient and home-based stroke rehabilitation. In *5th IEEE RAS/EMBS international conference on biomedical robotics and biomechatronics* IEEE, 438-443.
[39] Eyssel, F., 2017. An experimental psychological perspective on social robotics. *Robotics and Autonomous Systems 87*, 363-371.
[40] Wang, Y. and Quadflieg, S., 2015. In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social cognitive and affective neuroscience 10*, 11, 1515-1524.
[41] Robinson, H., Macdonald, B., and Broadbent, E., 2015. Physiological effects of a companion robot on blood pressure of older people in residential care facility: A pilot study. *Australasian journal on ageing 34*, 1, 27-32.

BiOpenBank information systems and its integration into the analysis of genetic predispositions in psychiatric disorders

Miha Moškon
Faculty of Computer and Information Science, University Ljubljana
+386-1-4798-217
miha.moskon@fri.uni-lj.si

Tadeja Režen
Centre for Functional Genomics and Bio-Chips, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana
+386-1-543-7592
tadeja.rezen@mf.uni-lj.si

Nataša Debeljak
Medical Centre for Molecular Biology, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana
+386-1-543-7645
natasa.debeljak@mf.uni-lj.si

Alja Videtič Paska
Medical Centre for Molecular Biology, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana
+386-1-543-7661
alja.videtic@mf.uni-lj.si

ABSTRACT
In this paper, we describe BiOpenBank, an open source information system devoted to the management of small biobanks. We describe its implementation, technologies that were used in the process of its development and its compatibility with recent standards, regulations and good practices, which present the state of the art in the field of biobanking as well as on the wider field of data management. We demonstrate the integration of BiOpenBank in the process of the analysis of genetic predispositions in psychiatric disorders.

Keywords
Biobank, data management, BIMS, GDPR, psychiatric disorders.

1. INTRODUCTION
Psychiatry is a medical specialty that has yet to establish clinically applicable biomarkers. In order to be able to provide well defined samples of patients and controls that would enable faster search for reliable, specific and sensitive biomarkers, biobanks dedicated to biological psychiatry have to be carefully planned. A more standardized approach could provide solid background for single biomarker research as well as for developing more holistic approaches oriented towards systems medicine, as psychiatric disorders are complex disorders and need to be treated as such.

The term biobank refers to a collection of biological, namely human, animal or plant, samples. Biobanks must handle different processes, such as collection, processing, storage and distribution of biological samples. Each of these must be compliant with a vast amount of requirements [1,2]. For example, storage conditions of samples need to be appropriate to maintain their integrity, access permissions need to be established and controlled, and audit trails of all the changes in the biobank need to be recorded. These requirements are defined by different standards, regulations and good practices, such as ISO 20387:2018 Standard [3], General Data Protection Regulation (GDPR) [4], Minimum Information About Biobank Data Sharing (MIABIS) [5,6], FDA 21 CFR Part 11 [7], ISBER best practices [8] and others. Imposed requirements are however hard to fulfill when data management is performed manually, i.e. without a designated biobanking information management system (BIMS). Recently, several commercially available BIMS platforms have been introduced. These present customizable solutions which are tailored to each user's demands. Moreover, their out-of-the-box compliance with the recent regulations and standards addresses all or at least most of the data protection and integrity requirements dictated either by legislature, common sense and/or good practices. The main problem of these solutions, however, is their cost, which makes them inaccessible for small, non-commercial research laboratories. Different open source BIMS solutions have already been reported recently, such as Advanced Tissues Management Application (ATiM), Baobab [9] and OpenSpecimen [10]. Even though these solutions might seem promising, they are still not fully compliant with the most recent requirements that are constantly being updated [11].

Herein, we describe our initiative to develop a comprehensive open source biobanking solution, which would be accessible to all, would be compliant with the most recent standards and regulations, and would allow straightforward customization. The described solution is a direct result of several projects which financed the collaboration within a vast interdisciplinary group of students, mentors and researchers. These projects were focused to the implementation of a general purpose BIMS in the domains of Genotyping in Alzheimer disease, research work with model organisms (mice), genetic predispositions in suicide victims, HCC biomarkers and genotyping in erythrocytosis. The group has been working on the implementation of the system from the initial specifications to its programming and testing in a laboratory environment. We describe the integration of the proposed solution in the process of the analysis of genetic predispositions in psychiatric disorders.

2. BIOBANKING STANDARDS AND REGULATIONS
Standards and regulations define the requirements that make the BIMS compliant with legislature, increase the safety and efficacy of the biobank, ensure the biobank integrity, as well as allow easier exchange of the samples and their corresponding data between different laboratories and research groups. Here, we only

briefly overview some of these standards and regulations, namely standard ISO 20387:2018 [3], GDPR [4], MIABIS 2.0 [5,6], FDA 21 CFR Part 11 [7] and ISBER best practices [8].

The International Organisation for Standardization (ISO) introduced the standard General Requirements for Biobanking ISO 20387:2018 mainly to promote the confidence in biobanking [3]. The standard aims to facilitate cooperation, fosters exchange and assist the harmonization of data and good practices among biobanks, researchers and other parties.

Harmonization of biobanks and facilitation of data exchange is also addressed by MIABIS [5,6]. MIABIS promotes the harmonization of biobanks by following the same set of Standard Operating Procedures (SOPs) and the same medical ontologies. Moreover, it defines the main biobank components and their data models, such as *Biobank*, *Sample collection* and *Study*.

GDPR is not specifically focused on the regulation of personal data within biobanks. However, biobanks operating within the European Union must comply with its requirements when handling personal data. These data can be collected only after the consent of the natural person has been acquired. Moreover, each individual has a right of the removal of consent, a right of erasure and a right to be forgotten. This means, that in the worst case, all the data belonging to this individual need to be removed from the biobank.

Part 11 of the Title 21 of the Code of Federal Regulations (FDA 21 CFR Part 11) is focused on the regulation on electronic records and their integrity [7]. One of its main requirements is that all the changes within the system should be automatically logged within the system's audit trail, which cannot be modified by anyone. Moreover, the system should implement role-privileged limited access.

Last but not least, the BIMS should follow good practices, which are defined by ISBER (International Society for Biological and Environmental Repositories) best practices [8]. These include topics, which are to some extent already addressed by other regulations and common practices.

3. IMPLEMENTATION

Biobank and BIMS requirements defined by recent standards, regulations and good practices served as a set of initial specifications of our system. These were updated with the functional requirements of the target users (collaborating laboratories). Functional requirements included the sample coding and decoding using QR codes, straightforward modularization of the user interface (principal investigator leading the study can choose among the modules, which will be present within the study), and use of the system on an arbitrary computational platform, such as personal computer, mobile phone or tablet, without any installation.

BiOpenBank was implemented as a web application running on a designated served. The system was designed according to the Model-View-Controller software architecture pattern. Laravel PHP framework was used to enhance the development process, and Laradock was used to configure the system within the docker environment. Data model was implemented within the PostgreSQL database. The source code of the BiOpenBank implementation is available at https://gitlab.com/biopenbank/biopenbank.

4. INTEGRATION OF BIOPENBANK AND ANALYIS OF GENETIC PREDISPOSITIONS IN PSYCHIATRIC DISORDERS

In order to perform reliable and reproducible research we have to be able to produce good quality and accessibility of samples and data. The problem of irreproducibility is very persistent as more than half of the errors stem from inappropriate manipulation of the samples during collection, preparation and storage of specimen [12]. There is namely an estimation that out of all preclinical studies 53.3% have errors, which means they are not reproducible. Among the most frequent errors are the ones concerning biological reagents and reference materials (36.1%), study design (27.6%), data analysis and reporting (25.5%) and laboratory protocols (10.8%) [13]. These errors could be mitigated with appropriate standardization of methods and procedures.

The search for reliable biomarkers is particularly intriguing in psychiatry, as there is currently no established laboratory testing that would aid physicians in their determination of the diagnosis, treatment protocol or monitoring of the patients [14]. For several years genetic testing has been an important research topic that might bring some specific and sensitive biomarker which could be used in clinical setting. Our most important research areas are suicidal behavior and depression, where we are looking for genetic markers.

In order to be able to standardize our procedure of sampling, storing and manipulation of the samples and data acquired during different projects in the field of psychiatric genetics, we studied standards, protocols and other published literature on the topic.

Based on the obtained information we identified the data that has to be included in the BIMS. According to MIABIS we first defined the data regarding the study, which describes the purpose of the research and designates the data and samples we are storing. Further on we defined the data we are going to collect about the study subjects, the sample handling and storage, isolated specimen, and the analyses performed.

The second step in the development of BIMS was study and preparation of the SOPs, which were prepared correspondingly to legislation and describe relevant processes in detail. They were prepared in accordance to the National Cancer Institute's Biorepositories and Biospecimen Research Branch [15], and are covering the following topics:

1. Informed consent.
2. Equipment monitoring, maintenance and repair.
3. Control of supplies used for biospecimen collection.
4. Biospecimen identification and labeling.
5. Methods for biospecimen collection and processing.
6. Sample storage and retrieval.
7. Shipping and receiving of samples.
8. Laboratory tests performed in-house including QC testing.
9. Biospecimen data collection and management.

10. Biosafety.
11. Training.
12. Security.

5. CONCLUSIONS

Safe and orderly storage of data and samples represents an important point in contemporary research. Particularly for smaller laboratories it represents an important challenge as they usually lack the resources to be able to use commercially available BIMS, while on the other hand open source BIMS are too general to enable efficient biobanking. In order to be able to participate in international projects, multicenter projects or just to be able to publish the results in established journals, the laboratories have to be able to collect, store, and manage numerous samples and their corresponding data. Development of an in-house BIMS encouraged our group to standardize the sample and data collection, storage, maintenance, and generation of SOPs which all importantly contributed to greater transparency of our work. Moreover, it improved our collaboration with clinical environment where regulations associated with biobanking are very demanding. It is expected that biobanking is going to undergo important changes in the upcoming years, which is in favor of their users, as in clinical setting only highly reproducible biomarkers can add value to the evolving fields of personalized and translational medicine.

6. ACKNOWLEDGMENTS

This work was partially supported by the projects "Establishment of open information system for the management of biological samples in the biobank repositories" and "Standardization of procedures for obtaining biological samples and information system for biobanks" co-financed by the Republic of Slovenia and the European Union under the European Social Fund. We would like to thank all the researchers, mentors and students collaborating in the process of BiOpenBank development, namely dr. Uršula Prosenc Zmrzljak, dr. Žiga Urlep, Kaja Blagotinšek Cokan, prof. dr. Damjana Rozman, Stane Moškon, Katarina Kouter, assist. prof. dr. Jurij Bon, prof. dr. Zvezdan Pirtošek, Jure Fabjan, Nejc Nadižar, Staš Hvala, Žiga Pušnik, Žiga Pintar, Marjeta Horvat, Živa Rejc, Luka Toni, Roman Komac, Matevž Fabjančič, Andraž Povše, Eva Drnovšek, Zala Gluhić, Fran Krstanović, Jernej Janež, Filip Grčar, Gašper Vrhovnik, Nermin Jukan, Laura Bohinc, Natalija Pucihar, Julija Lazarovič, Andrej Gorjan, Tilen Burjek, Kity Požek, Tina Turel, Sara Tekavec, Nejka Kotnik, Vid Rotvejn Pajič and Nataša Vodopivec.

7. REFERENCES

[1] Dollé, L. and Bekaert, S., 2019. High-quality biobanks: pivotal assets for reproducibility of OMICS-data in biomedical translational research. *Proteomics.* 1800485. DOI= https://doi.org/10.1002/pmic.201800485.

[2] Janzen, W., Admirand, E., Andrews, J., Boeckeler, M., Jayakody, C., Majer, C., Porwal, G., Sana, S., Unkuri, S. and Zaayenga, A., 2019. Establishing and Maintaining a Robust Sample Management System. *SLAS TECHNOLOGY: Translating Life Sciences Innovation.* 24, 3, 256-268. DOI= https://doi.org/10.1177/2472630319834471.

[3] ISO. *Biotechnology – Biobanking – General requirements for biobanking. ISO 20387:2018(E)*, International Organization for Standardization, Geneva, Switzerland, August 2018.

[4] EU Commission. 2018. *EU GDPR Portal*. Accessible at https://eugdpr.org/.

[5] Norlin, L., Fransson, M.N., Eriksson, M., Merino-Martinez, R., Anderberg, M., Kurtovic, S. and Litton, J.E., 2012. A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreservation and biobanking.* 10, 4, 343-348. DOI= https://doi.org/10.1089/bio.2012.0003.

[6] Merino-Martinez, R., Norlin, L., van Enckevort, D., Anton, G., Schuffenhauer, S., Silander, K., Mook, L., Holub, P., Bild, R., Swertz, M. and Litton, J.E., 2016. Toward global biobank integration by implementation of the minimum information about biobank data sharing (MIABIS 2.0 Core). *Biopreservation and biobanking.* 14, 4, 298-306. DOI= https://doi.org/10.1089/bio.2015.0070.

[7] U.S. Food & Drug Administration. 2003. *Part 11, Electronic Records; Electronic Signatures - Scope and Application*. Accessible at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application.

[8] Campbell L.D., Astrin J.J., DeSouza Y., Giri, J., Patel A.A., Rawley-Payne M., Rush A. and Sieffert N. 2018. *The 2018 Revision of the ISBER Best Practices: Summary of Changes and the Editorial Team's Development Process*. 16, 1, 3-6. DOI= https://doi.org/10.1089/bio.2018.0001.

[9] Bendou, H., Sizani, L., Reid, T., Swanepoel, C., Ademuyiwa, T., Merino-Martinez, R., Meuller, H., Abayomi, A. and Christoffels, A., 2017. Baobab Laboratory Information Management System: Development of an Open-Source Laboratory Information Management System for Biobanking. *Biopreservation and biobanking.* 15, 2, 116-120. DOI= https://doi.org/10.1089/bio.2017.0014.

[10] McIntosh, L.D., Sharma, M.K., Mulvihill, D., Gupta, S., Juehne, A., George, B., Khot, S.B., Kaushal, A., Watson, M.A. and Nagarajan, R., 2015. caTissue suite to OpenSpecimen: Developing an extensible, open source, web-based biobanking management system. *Journal of biomedical informatics.* 57, 456-464. DOI= https://doi.org/10.1016/j.jbi.2015.08.020.

[11] Vodopivec, N. 2019. Integration of standards and guidelines into the open source information management systems for biobanks. Master Thesis. Université Grenoble Alpes.

[12] Litton, J.E., 2017. Reproducibility and reliability. *Biobanks Europe.* 6, 2-3.

[13] Freedman L.P., Cockburn I.M., Simcoe T.S. 2015. The Economics of Reproducibility in Preclinical Research. *PLOS Biology.* 13, 6, e1002165. DOI= https://doi.org/10.1371/journal.pbio.100262.

[14] Venkatasubramanian G. and Keshavan M.S. 2016. Biomarkers in Psychiatry - A Critique. *Annals of Neurosciences*. 23, 1, 3–5. DOI= https://doi.org/10.1159/000443549.

[15] National Cancer Institute. 2017. *NCI Best Practices for Biospecimen Resources*. Accessible at https://biospecimens.cancer.gov/bestpractices/2016-NCIBestPractices.pdf

Comment sentiment associations with linguistic features of educational video content

Lenart Motnikar
Faulty of Education,
University of Ljubljana,
Slovenia
lenart.motnikar@gmail.com

ABSTRACT

As people spend an increasing amount of time on social media, researchers are motivated to study the newly emerging communities and the interpersonal relationships within them. This study examines one such relationship, namely between the audiences of educational videos and its presenters. A dataset of sentiment-labeled comments from TEDx and TED-Ed YouTube videos was extended to include linguistic features of video content. It was revealed that the features significantly varied between animations and presentations, and in the latter case, the speakers' genders. A correlation analysis showed that sentiment depended on a number of features, where the most notable observations included associations between negative sentiment and negative emotional content, and between positive sentiment and (first person singular) personal pronouns.

Keywords

TED Talk, YouTube, comments, LIWC, sentiment analysis

1. INTRODUCTION

As social media platforms like YouTube became so prevalent in our daily lives [1], offering opportunities for interaction with wide audiences, educators and scholars are often motivated to participate with their own content [2]. The interactions on these platforms, however, are not always civil, and are frequently characterized by unwanted behavior [3]. In order to foster better online communities, recent research has focused on understanding contentious individuals and studying the effects of various design and moderation measures [4, 5]. Research has also suggested that individuals sharing content online mind the potential reactions of their audience and, motivated by not being badly perceived, adapt their behavior accordingly [6]. Little research has, however, been done on the specifics of these behavioral measures, or their effectiveness in terms of influencing the audience. A study examining vloggers, for example, found that they use a distinctive viewer-oriented speaking style, often characterized by explicit or implicit encouragements of desired behaviors (e.g. commenting, subscribing) [7]. Building upon these observations, this study, using a quantitative approach, explores potential ways for content creators to influence their audiences' behavior. By applying methods and theory previously unused in such a setting it explores associations between the language used in educational videos and the sentiments expressed in the comments, opening opportunities for future inquiries into the dynamics between individuals and large online audiences.

1.1 Lexical inquiry and word count (LIWC)

For linguistic analysis, the Linguistic Inquiry and Word Count (LIWC) program was used [8]. LIWC is a text analysis software which, using a predefined dictionary, measures the frequency of words across a variety of categories relating to grammar and psychological processes, and rates the text's manifestations of four underlying psychological dimensions - analytical thinking, authenticity, clout (expression of social status) and emotional tone. In the last two decades LIWC has become the most popular tool for automated text analysis in socio-psychological studies, as it helped illuminate how a person's choice of words reflects their mental states (for a review, see [9]). One of the most notable revelations stemming from LIWC research was the importance of function words in human social dynamics. Personal pronouns were shown to be particularly revealing as they, by conveying information about attentional focus, let us know how people relate to themselves and others, disclosing details ranging from one's social status to their emotional states.

1.2 Sentiment analysis

The research field of sentiment analysis or opinion mining aims to capture the public's feelings about various entities, be it products, people or ideas [10]. Due to the availability of a wide variety of tools and data, a significant portion of the field deals with the analysis of texts gathered from social media. The sentiment in this study was assessed with the SentiStrength [11] tool, which, using a lexical approach, identifies sentiment-related tokens and scores social web texts on a dual positive and negative scale.

1.2.1 Comment sentiment on TED YouTube videos.

The current study builds upon a dataset compiled by Veletsianos et al. [12]. The authors collected English-speaking educational YouTube videos posted on TEDx Talks and TED-Ed channels and investigated how presenter gender, video format and comment threading effect the sentiment expressed in the comments. They observed that presentations with female presenters, relative to those with male, exhibited greater polarity in positive and negative sentiment, and that animated videos were more neutral than presentations. These differences not only held for comments directed toward the video, but replies to the comments as well. The study also examined the relationship between sentiment and video topic by analyzing description and title keywords, and found that some topics exhibit more positive (e.g. beauty) and others more negative (e.g. cancer) sentiment.

2. METHOD

A modified »YouTube TED Talk Comment Sentiment Data« dataset [13] was used. The dataset contained positive (1 to 5) and negative (-1 to -5) sentiment scores of comments from 665 videos, information about whether the video was an animation or a presentation, and in the latter case, the information about presenter's gender. In this study, the dataset was extended to include LIWC scores of video subtitles. The subtitles were assessed using the LIWC2015 dictionary, scoring each subtitle track across 93 linguistic categories. As not every subtitle track featured all of the categories, occurrences where the score of a category equaled zero were ignored in the analysis.

Videos that did not have English subtitles available were excluded from the dataset (n = 57), reducing the sample size of videos and comments by 8.6% and 6.7%, respectively. Additionally, the analysis only included first-level comments representing 50% of the sample. Because comments on YouTube come in two general forms, posted directly under the video or as a reply to another comment, this study followed the interpretation that replies are directed towards other comments rather than the video itself.

Table 1: Descriptive statistics of videos and comments

Format/ gender	Videos *n*	Comment *n*	Comment *n* M	Comment *n* SD
Female	66	38572	584.42	782.40
Male	130	89642	689.55	1575.45
Animation	412	197385	479.09	873.51
	608	325599	535.52	1056.97

While the removal of videos minimally affected the reported differences between video formats and presenters, the exclusion of replies significantly increased both positive and negative average sentiment. The general trend that videos with female presenters exhibited greater polarity and that animations were the most neutral, however, still remained.

Table 2: Sentiment differences of comments by format and gender

Format/ gender	Positivity		Negativity	
	M	SD	M	SD
Female Speaker	2.16	0.98	-1.72	1.06
Male Speaker	1.96	0.95	-1.63	0.98
Animation	1.60	0.78	-1.62	0.94

Each video then received two aggregated sentiment scores by separately averaging the positive and negative sentiment of all its comments.

Table 3: Differences of aggregated sentiment scores by format and gender

Format/ gender	Positivity M	Negativity SD	Positivity M	Negativity SD
Female	2.23	0.21	-1.71	0.26
Male	2.02	0.25	-1.61	0.31
Animation	1.63	0.18	-1.58	0.27

This further increased the average positivity and negativity, reflecting the otherwise statistically insignificant trend that sentiment averages decrease as the number of comments on a video increases.

3. RESULTS

The data was tested for differences in LIWC scores between video formats and presenter gender. The Wilcoxon rank sum test revealed that the video formats significantly ($p < 0.01$) differed in 70 and genders in 26 of the 93 LIWC2015 categories. Differences in summary variables and language metrics showed that animations were more analytical and used longer words and sentences, whereas the presentations had a greater word count and exhibited more clout, authenticity and emotional tone. Similar differences could be observed between the genders, where videos with male presenters exhibited greater analytical thinking and those with female presenters more authenticity. Numerous differences in categories relating to style and content were also observed, a selection of which is shown in Table 4.

Table 4: Word category prevalence by format and gender

	Animation	–	Talk
Female	anxiety, body	*negative emotion, sadness, female referents, feeling, health*	pronouns, 1st person singular, regular verbs, conjunctions, negations, *affect, certainty*
–	prepositions, adjectives, comparatives, *death, anger, seeing, sexuality, ingesting, relativity, space, religion, friends, swearing*	3rd person, *tentativeness, differentiation, assent, home*	1st person plural, 2nd person, auxiliary verbs, adverbs, interrogatives, *positive emotion, social processes, insight, discrepancies, hearing, time orientation, drives, motion, work*
Male	articles, quantifiers, numbers	*money*	*informal speech, leisure*

Note. Content categories are presented in italic

Across the five (sub)samples, correlating positive and negative aggregated sentiments with LIWC scores revealed 302 significant ($p < 0.05$) correlations, of which 83 were stronger than $|r| = 0.3$. Because the correlations covered a large majority of the LIWC2015 categories, only the categories exhibiting correlations above $|r| = 0.3$ in at least two sentiment-sample pairings are reported in Table 5. In the sample containing all videos, correlations with three out of four summary variables could be observed. Positive sentiment was positively associated with authenticity and inversely with analytic thinking, while emotional tone positively correlated with both positive and negative sentiment (note that negative sentiment was represented by a value between -1 and -5). The association between emotional tone and negative sentiment, however, remained in all samples. Significant correlations with language metrics could also be observed. The percentage of words longer than six letters exhibited a general inverse correlation with negative sentiment,

and in the case of videos with female speakers, positive sentiment as well.

was related more to content whereas positive sentiment to style and grammar, especially (first person singular) personal pronouns.

Table 5: Correlations between aggregated sentiments and LIWC categories

		Positive sentiment					Negative sentiment				
		All videos	Presentations			Animated videos	All videos	Presentations			Animated videos
	LIWC categories		All	Male	Female			All	Male	Female	
Summary variables	Analytic thinking	-.62***	-.24***	-.08	-.33**	-.08	.06	-.19**	-.23**	-.24*	.08
	Authenticity	.28***	.31***	.16	.45***	-.12*	.04	.01	.05	.06	.11*
	Tone	.27***	-.06	-.05	.02	.09	.28***	.44***	.40***	.49***	.30***
Language metrics	Words >6 letters	-.40***	-.17*	-.05	-.42***	.05	-.14***	-.23**	-.22*	-.30*	-.24***
	Dictionary words	.56***	.37***	.29**	.36**	.00	-.13**	.06	.09	.14	-.14**
Style and grammar	Function words	.57***	.32***	.20*	.35**	-.05	-.00	.16*	.15	.35**	.07
	Total pronouns	.66***	.34***	.20*	.44***	.11*	-.05	.16*	.16	.34**	-.01
	Personal pronouns	.67***	.45***	.29***	.56***	.12*	-.07	.11	.13	.24	-.06
	1st person singular	.63***	.40***	.19*	.60***	.23*	-.09	-.01	.04	.04	-.04
	Articles	-.46***	-.28***	-.09	-.40***	-.10	.15***	-.04	-.13	.01	.18***
	Regular verbs	.55***	.16*	-.02	.36**	.00	-.00	.19**	.22*	.25*	.05
	Quantifiers	-.23***	-.31***	-.17	-.42***	-.03	.15***	.02	-.01	-.06	.17***
Content	Affect words	.34***	.18*	.11	.18	.24***	-.39***	-.17*	-.10	-.30*	-.47***
	Negative emotion	.10*	.21**	.16	.12	.14**	-.53***	-.44***	-.35***	-.61***	-.58***
	Anger	-.08	.029	.07	-.01	.08	-.28***	-.19*	-.17	-.35**	-.35***
	Sadness	.06	.20**	.15	.16	.15*	-.36***	-.42***	-.29**	-.66***	-.36***
	Biological processes	-.04	.16*	.14	-.01	.04	-.20***	-.33***	-.33***	-.27*	-.19***
	Health	-.03	.15*	.17	-.03	-.05	-.34***	-.36***	-.37***	-.31*	-.35***
	Past focus	.35***	.31***	.23**	.41***	.05	-.07	-.05	-.06	.05	-.03
	Death	-.25***	.01	.07	.11	-.08	-.19***	-.44***	-.53***	-.22	-.18**

Note. For visualization purposes, the significant correlations are colored with a grey-to-black gradient, representing their strength.
*p<0.05, **p<0.01, ***p<0.001

The presentation subsamples also exhibited correlations between positive sentiment and the percentage of words caught by the dictionary. Regarding style and grammar, positive sentiment was associated with function words, particularly (first person singular) personal pronouns. In the female presenter subsample, associations with positive sentiment were observed between regular verbs, quantifiers and articles, while negative sentiment positively correlated with the percentages of function words, pronouns and verbs. Contentwise, a majority of significant correlations was with negative sentiment, most of which were inverse and related to negative affective processes like anger and sadness, or concerns like health and death. Positive sentiment exhibited fewer and weaker content related associations, except in the case of presentations expressing a greater focus on the past.

4. DISCUSSION AND CONCLUSIONS

An important caveat before delving into interpretations is that the videos included in this study had different audiences. In fact, more than 90% of commenters only commented on one or two videos, as different topics and formats invite different profiles of people. While this does not change the overall experience for the comment reader, it should be noted that the results would likely differ with a constant or randomized audience.

Nevertheless, the analysis returned some interesting results. A general pattern was observed, showing that negative sentiment

The association with content is not that surprising as it can at least partially be attributed to video topic, as has been reported in the original study. Additionally, the emotion tokens SentiStrength and LIWC used for analysis overlap to some degree. This explanation also holds for the association with emotional tone, as it merely combines the words from emotion categories.

From a socio-psychological perspective the association between positive sentiment and style is more intriguing. While the importance of function words in human social dynamics is well documented, it has so far been limited to studies of smaller groups of people, like couples or teams [14]. This is the first time that a reaction of a larger audience has been associated with a speaker's pronoun use. What this observation means in terms of social psychology is less clear. It should be noted that sentiment, as it was assessed here, is a theoretically unsound construct and a particularly crude measure of emotion (for a critique, see [15]). It only measures emotion on a dual positive/negative scale, and does not differentiate between the nuances of human emotional experience and expression. For example, on a video discussing suicide, a comment personally attacking the speaker might receive the same sentiment score as one where the commenter shares their experience with depression. The motivations for these behaviors are vastly different, as are the readers' reactions. For this reason, one should be careful when interpreting sentiment and take into account the variety of factors contributing to its manifestation.

These limitations considered, the observed associations still encompass some psycholinguistic information about speaker-audience interaction, and call for a deeper inquiry into the topic.

A question that still remains is why the sentiments were differently associated with content and style in the first place. The observation may reveal information about the social aspects of emotion processing. If we only focus on the clearest examples, negative emotion and first person singular, a general explanation could be that the former evokes more sympathy whereas the latter, which entails more self-focus, evokes cheer.

Results also suggest a relationship between sentiment and language metrics, specifically the percentages of words longer than six letters and that of words caught by the dictionary. As the dictionary encompasses some 6000 words and stems in common use, this observation might indicate a relation to the simplicity or commonality of language used in the video. This could be interpreted in a way that people prefer simpler language, or that the use of more complex language encourages more sentiment-neutral conversation.

Lastly, the results shed light on the originally reported gender and format differences in sentiment. The groups varied in content and style, which might entail that some of the primarily observed discrepancies were due to the differences in topics the content makers chose, or the ways in which they were expressed. This considered, this explanation likely accounts only for a portion of the difference as there was still notable variation in correlation strengths between the samples, with the female subsample exhibiting the strongest correlations in most categories. For example, in the female subsample, but not the other two, positive sentiment exhibited an inverse correlation with articles and quantifiers. While this could still be due to the chosen topics, or some other confounding factor, another explanation for the phenomenon may lay in the fact that these words are mostly used in conjunction with concrete nouns, indicating a relation to concreteness or abstractness of a presentation. Why this relation would be only specific to female presenters, remains an open question.

Taken together, this study was mostly exploratory in nature, providing more avenues for research than solid findings. In order to thoroughly answer the questions emerged, future research should use more sound measures of behavior and mental states, as well as look into different communities and platforms where similar interpersonal interactions take place.

5. REFERENCES

[1] Smith A. and Anderson M. 2018. Social Media Use in 2018. *Pew Research Center* (March, 2018).

[2] Gruzd, A., Haythornthwaite, C., Paulin, D., Gilbert, S. and del Valle, M. 2016. Uses and Gratifications factors for social media use in teaching: Instructors' perspectives. *New Media & Society*, 20, 2 (February, 2018), 475-494.

[3] Duggan M. 2017. Online Harassment 2017. *Pew Research Center* (July, 2017).

[4] Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C. and Leskovec, J. 2017. Anyone Can Become a Troll. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, February 25 - March 01, 2017). CSCW '17. ACM, New York, NY, 1217-1230.

[5] Cheng, J., Dansecu-Nisulescu-Mizil, C. and Leskovec, J. 2015. Antisocial behavior in online discussion communities. In *Ninth International AAAI Conference on Web and Social Media* (Oxford, UK, May 26 - 29, 2015). ICWSM '15. AAAI, Palo Alto, California, 61-70.

[6] Barasch, A. and Berger, J. 2014. Broadcasting and Narrowcasting: How Audience Size Affects What People Share. *Journal of Marketing Research* 51, 3 (June, 2014), 286-299.

[7] Frobenius, M. 2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics* 72, (October, 2014), 59-72.

[8] Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. 2015. *The development and psychometric properties of LIWC2015*. University of Texas at Austin, Austin, TX.

[9] Tausczik, Y. R. and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29, 1 (March, 2010), 24-54.

[10] Cambria, E. 2016. Affective Computing and Sentiment Analysis. *IEEE Intelligent Systems* 31, 2 (March, 2016), 102-107.

[11] Thelwall, M. 2017. Heart and soul: Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions: Collective emotions in cyberspace*, J. Holyst, ED. Springer, Berlin, 119-134.

[12] Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T. and Lowenthal, P. 2018. Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. *PLOS ONE* 13, 6 (2018), e0197331.

[13] Kimmons, R. 2019. YouTube TED Talk Comment Sentiment Data. *BYU ScholarsArchive*, https://scholarsarchive.byu.edu/data/3.

[14] Kacewicz, E., Pennebaker, J., Davis, M., Jeon, M. and Graesser, A. 2013. Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology* 33, 2 (March, 2014), 125-143.

[15] Paolillo, J. 2019. Against 'Sentiment'. In *Proceedings of the 10th International Conference on Social Media and Society* (Toronto, ON, Canada, July 19 - 21. 2019). SMSociety '17. ACM, New York, NY, 41-48.

Primerjava kognitivnih sposobnosti igralcev akcijskih videoiger in neigralcev videoiger

Comparison of cognitive skills between action video game players and non-gamers

Neža Podlogar
Filozofska fakulteta
Univerza v Ljubljani
Aškerčeva 2
1000 Ljubljana
neza.podlogar@gmail.com

Anja Podlesek
Filozofska fakulteta
Univerza v Ljubljani
Aškerčeva 2
1000 Ljubljana
anja.podlesek@ff.uni-lj.si

IZVLEČEK

V raziskavi na slovenskem vzorcu preverjamo povezavo med igranjem akcijskih videoiger (AVI) in sposobnostjo mentalne rotacije, sledenja več objektom in preklapljanja med nalogami. Rezultati so pokazali, da je igranje AVI pomemben napovednik sposobnosti mentalne rotacije in krajših reakcijskih časov pri nalogi preklapljanja. Čeprav so igralci AVI hitreje preklapljali med nalogami pri vseh pogojih, pa se skupini nista razlikovali v stroških preklapljanja, ki so glavni pokazatelj kognitivne fleksibilnosti. Povezava z obsegom pozornosti ni bila jasna, statistično pomembne razlike med igralci in neigralci so bile opazne po izključitvi enega izstopajočega udeleženca. Rezultati nakazujejo, da je igranje AVI pozitivno povezano z določenimi kognitivnimi sposobnostmi, vendar zahtevajo nadaljnje presečne in eksperimentalne študije, ki bi dale več informacij o vzrokih in mehanizmih izboljšanja sposobnosti.

Ključne besede

Akcijske videoigre, prostorska sposobnost, mentalna rotacija, obseg pozornosti, preklapljanje med nalogami

ABSTRACT

Our research examines the connection between playing action video games (AVG) and the ability to mentally rotate objects, track multiple objects, and switch between tasks. The results show that playing AVG is an important predictor of mental rotation ability and faster reaction times in task switching. Even though AVG players switched between tasks more quickly than non-gamers in all conditions, the groups did not differ in the switching cost, which is a major indicator of cognitive flexibility. The effect on attention span was not as clear; statistically significant differences between action gamers and non-gamers were noticeable after excluding one participant. The results indicate that playing AVG can have positive effects on certain cognitive functions, but require further cross-sectional and experimental studies to provide more information on the causes and mechanisms of cognitive abilities improvement.

Keywords

Action video games, spatial ability, mental rotation, attention span, task switching

1. UVOD

Videoigre so dandanes povsod med nami, napovedi pa kažejo, da bodo v prihodnosti vse bolj razširjene. Najbolj popularna zvrst je akcijska, ki je tudi najbolj zanimiva z vidika kognitivne psihologije. Raziskovanje, kako igranje lahko vpliva oziroma, ali je povezano s kognicijo, je v porastu, kljub temu pa v Sloveniji to še ni raziskano področje. Čeprav so bile AVI primarno izdelane za zabavo in prosti čas, vse večje število raziskav kaže, da ima igranje te zvrsti pozitiven učinek na širok spekter zaznavnih in kognitivnih sposobnosti. Dve nedavni metaanalizi [1, 6] kažeta, da se igranje AVI povezuje s prostorsko kognicijo, pozornostjo, vodeno od zgoraj navzdol, izvršilnimi funkcijami in verbalno kognicijo. Obstaja tudi empirična podpora za vzročne učinke na področju prostorske kognicije in pozornosti. Kljub temu povezave med igranjem in kognitivnimi sposobnostmi še niso dobro raziskane; vzorci v raziskavah so pogosto majhni, definicija akcijske zvrsti pa nenatančna, zaradi česar prihaja do neprimernega uvrščanja nekaterih igralcev v skupino akcijskih. Prav tako omenjeni metaanalizi poročata o različnih velikostih učinkov, zato so za zanesljivejše zaključke potrebne dodatne študije, ki bi se izognile omenjenim pomanjkljivostim.

V ta namen smo razvili spletno računalniško testiranje, ki je vsebovalo tri kognitivne teste. Osredotočili smo se na sposobnost mentalne rotacije, ki je del prostorskih sposobnosti, obseg pozornosti in sposobnost preklapljanja med nalogami, ki je del izvršilnih funkcij.

2. METODOLOGIJA

2.1. Udeleženci

Vzorčenje je bilo neslučajnostno, saj smo načrtno iskali igralce in neigralce videoiger. Sodelovalo je 452 posameznikov, vendar nekateri niso zaključili meritev ali niso ustrezali kriterijem. Končni vzorec je vključeval 163 udeležencev (starih 18–37 let), od tega 82 igralcev (70 moških, 12 žensk) in 81 neigralcev (37 moških, 44 žensk).

2.2. Pripomočki

2.2.1. Vprašalnik o igranju videoiger

Za razvrstitev v skupino igralcev in neigralcev smo uporabili vprašalnik o igranju videoiger.[1] Udeleženec je za dva časovna sklopa (v preteklem letu in pred preteklim letom) in vsako od sedmih kategorij videoiger izpolnil, kako dober je v tej kategoriji, pogostost igranja in katere videoigre je igral. Posameznik je bil uvrščen v skupino igralcev AVI, če je v zadnjih 12 mesecih igral AVI vsaj 6 ur na teden, pri čemer drugih zvrsti ni igral pogosto. Za uvrstitev v skupino neigralcev je moral poročati, da AVI v preteklem letu ni igral, prav tako ni smel imeti veliko izkušenj z igranjem drugih zvrsti.

2.2.2. Test mentalne rotacije (MRT)

Za preverjanje sposobnosti mentalne rotacije smo uporabili test mentalne rotacije (Mental Rotations Test – MRT [7]). Sestavljen je iz dveh delov, vsak del obsega 10 nalog. Vsaka naloga je sestavljena iz osnovnega objekta na levi in štirih alternativ na desni. Posameznik mora izmed štirih možnosti izbrati dve, ki sta enaki osnovnemu objektu. Edina razlika med osnovnim objektom in pravilnim odgovorom je v zornem kotu oz. rotaciji. Pravilna odgovora sta pri vsaki nalogi samo dva. Reševanje je časovno omejeno na 6 minut (3 minute za vsak del). Maksimalno možno število točk je 40. Dve točki dodelimo za oba pravilno izbrana odgovora, 1 točko, če je izbran le en pravilen odgovor, in 0 točk, če sta izbrana pravilen in nepravilen odgovor ali le nepravilni odgovori.

2.2.3. Test sledenja več objektom (MOT)

Test sledenja več objektom (Multiple Object Tracking – MOT [8]) smo uporabili kot mero obsega vidne pozornosti. Udeleženec mora vso pozornost usmeriti v naključno premikajočih se 16 rumenih krogov. Po dveh sekundah se določeno število krogov (1–5) obarva modro in tem mora slediti. Po štirih sekundah sledenja se vsi krogi obarvajo nazaj v prvotno rumeno barvo. Nato se le enega izmed 16 krogov izpostavi in udeleženec mora odgovoriti, ali je to dražljaj, kateremu je moral slediti (modro obarvan), ali ne. Test vsebuje 6 nalog za vajo, nato sledi 45 poskusov, razdeljenih v tri sklope (vsak sklop ima 15 poskusov).

2.2.4. Test preklapljanja Switcher

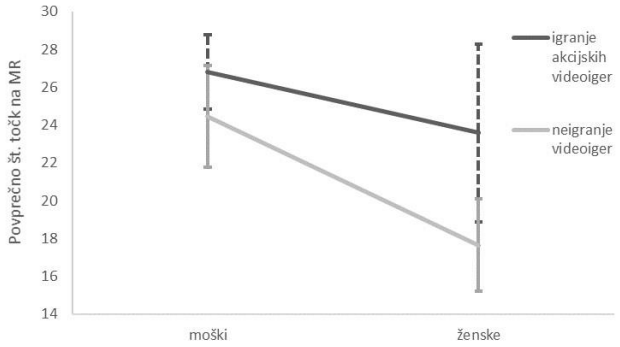
S testom preklapljanja (The PEBL Switcher Task [5]) smo merili kognitivno fleksibilnost oz. sposobnost fleksibilnega preklapljanja med nalogami z različnimi pravili. Na zaslonu je naključno razporejenih 10 dražljajev, ki se razlikujejo po barvi, obliki in črki. Na začetku vsakega preizkusa je en dražljaj obkrožen in na vrhu zaslona napisano pravilo, kateremu mora udeleženec slediti tako, da izbere naslednji ustrezni dražljaj. Tri pravila so »barva«, »oblika« in »črka«. Če je pravilo npr. »oblika«, mora udeleženec poiskati dražljaj, ki je enake oblike kot dražljaj, ki je takrat obkrožen. Test je razdeljen na tri stopnje preklapljanja; pri prvi se v enakem vrstnem redu izmenjujeta dve pravili, pri drugi se v enakem vrstnem redu izmenjujejo tri pravila, pri tretji pa se naključno izmenjujejo tri pravila. Vsaka stopnja preklapljanja vsebuje tri preizkuse z devetimi preklopi. Meril se je reakcijski čas in število napak.

2.3. Analiza podatkov

Pri testu mentalne rotacije smo podatke analizirali z dvosmerno ANOVO za neponovljene meritve, kot faktorja smo določili spol in (ne)igranje videoiger. Pri testu sledenja objektom smo rezultate analizirali z 2 (igralci/neigralci – neponovljene meritve) x 4 (dva, tri, štiri in pet objektov sledenja – ponovljene meritve) ANOVO. Opravka smo imeli z deleži od 0 do 1, vendar je bilo meritev več, vrednosti pa se niso gibale le okoli 0 ali 1 (z izjemo sledenja enemu objektu, ki je bil iz analize izločen), zato je bila uporaba ANOVE smiselna. Pri testu preklapljanja smo rezultate analizirali z 2 (igralci/neigralci – neponovljene meritve) x 3 (prva, druga, tretja stopnja preklapljanja – ponovljene meritve) ANOVO. Če je bila pri Mauchlyjevem testu stopnja nesferičnosti statistično pomembna, smo uporabili Huynh-Feldtov popravek prostostnih stopenj. Za preverjanje razlik v časih pri različnih stopnjah preklapljanja smo uporabili *t*-test za dva neodvisna vzorca, za preverjanje razlik v številu napak pa neparametrični test Mann-Whitney *U*, postopek Monte Carlo.

3. REZULTATI

Igralci (M = 26, SD = 8) so na testu mentalne rotacije v povprečju dosegli 5 točk višji rezultat od neigralcev (M = 21, SD = 9). Ta razlika je bila statistično značilna, $F(1, 158) = 6{,}86$, $p = {,}010$, $\eta_p^2 = 0{,}04$. Z rezultati mentalne rotacije je bil pomembno povezan spol, $F(1, 158) = 10{,}07$, $p = {,}002$, $\eta_p^2 = 0{,}06$, in sicer so moški dosegli višje rezultate (M = 26, SD = 8) kot ženske (M = 19, SD = 9). Interakcija med učinkom igranja AVI in spolom ni bila statistično značilna, $F(1, 158) = 1{,}30$, $p = {,}255$, $\eta_p^2 = 0{,}008$. To pomeni, da je učinek igranja AVI na sposobnost mentalne rotacije podoben pri moških in ženskah. Rezultate na testu ponazarja slika 1.



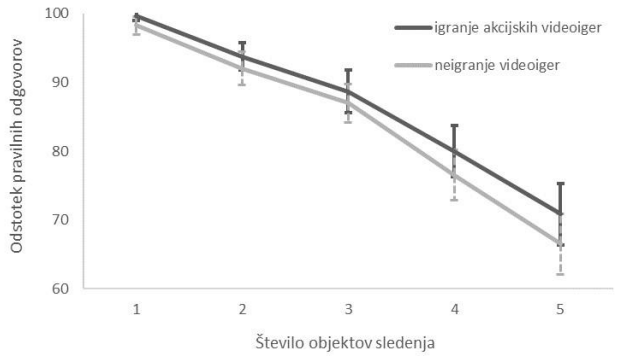
Slika 1: Povprečni dosežki s 95 % IZ na testu mentalne rotacije glede na spol in igranje videoiger

Pri testu sledenja objektom so imeli igralci v povprečju 85,1 % pravilnih odgovorov (SD = 7,5 %), neigralci pa 82,5 % (SD = 7,9 %). Razlika med skupinama igralcev in neigralcev ni bila statistično pomembna, $F(1, 126) = 3{,}37$, $p = {,}069$, $\eta_p^2 = 0{,}026$. Na rezultate je pomembno vplival učinek števila objektov sledenja, $F(2{,}63;\ 331{,}27) = 88{,}8$, $p < {,}001$, $\eta_p^2 = 0{,}413$. Več kot je bilo objektov, ki jim je moral posameznik slediti, nižji je bil povprečni delež pravilnih odgovorov. Interakcija med igranjem videoiger in številom sledenih objektov ni bila statistično pomembna, $F(2{,}63;\ 331{,}27) = 0{,}338$, $p = {,}771$, $\eta_p^2 = 0{,}003$.

En udeleženec iz skupine igralcev je zelo odstopal od povprečja svoje skupine (za –3,07 SD), zato smo ga izključili in analizo rezultatov ponovili. Pri tem je ANOVA pokazala statistično pomembno razliko med skupinama, $F(1, 125) = 4{,}80$, $p = {,}037$, $\eta_p^2 = 0{,}037$. Na rezultate je pomembno vplival učinek števila objektov sledenja, $F(2{,}60;\ 324{,}50) = 88{,}3$, $p < {,}001$, $\eta_p^2 = 0{,}414$. Interakcija med igranjem videoiger in številom sledenih objektov ni bila

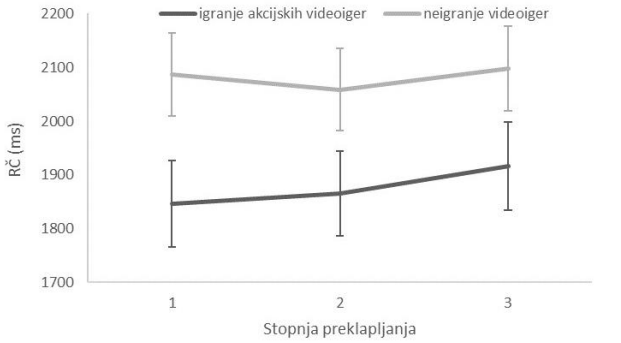
[1] Vprašalnik in kriteriji so bili prirejeni po https://osf.io/t72vp/.

statistično pomembna, $F(2{,}60;\ 324{,}50) = 0{,}395$, $p = {,}728$, $\eta_p^2 = 0{,}003$ (glej sliko 2).



Slika 2: Povprečni odstotki pravilnih odgovorov igralcev in neigralcev s 95 % IZ pri 1–5 objektih sledenja

Iz slike 3 je razvidno, da so imeli igralci pri testu preklapljanja v vseh pogojih naloge krajše reakcijske čase od neigralcev. Razlika med skupinama je bila statistično pomembna, $F(1, 141) = 15{,}679$, $p < {,}001$, $\eta_p^2 = 0{,}10$. Na hitrost reagiranja je pomembno vplivala tudi stopnja preklapljanja, $F(2, 282) = 3{,}230$, $p = {,}042$, $\eta_p^2 = 0{,}022$. Reakcijski časi so v povprečju naraščali s stopnjami preklapljanja. Najboljši pokazatelj kognitivne fleksibilnosti pa je primerjava različnih stopenj preklapljanja, ki nam poda informacijo o stroških preklapljanja. Interakcija med igranjem videoiger in stopnjami preklapljanja ni bila statistično značilna, $F(2, 282) = 1{,}240$, $p = {,}291$, $\eta_p^2 = 0{,}009$, kar pomeni, da je bil učinek igranja videoiger podoben pri vseh stopnjah preklapljanja. Igralci in neigralci se niso statistično pomembno razlikovali v stroških preklapljanja. Razlika v reševanju nalog s kompleksnejšimi preklopi in manj kompleksnimi preklopi je bila v obeh skupinah podobna. Pri primerjavi razlik v reakcijskih časih preklapljanja med tremi in dvema praviloma v konsistentnem zaporedju (2. in 1. stopnja preklapljanja) se igralci in neigralci niso pomembno razlikovali, $t(141) = 1{,}19$, $p = {,}237$, enako je pokazala primerjava razlik v reakcijskih časih naključnega in konsistentnega preklapljanja med tremi pravili (3. in 2. stopnja preklapljanja), $t(142) = 0{,}298$, $p = {,}766$.



Slika 3: Povprečni reakcijski časi igralcev in neigralcev s 95 % IZ pri treh različnih stopnjah preklapljanja

Kljub temu, da so bili igralci pri vseh pogojih testa preklapljanja hitrejši, pa pri tem niso naredili pomembno več napak (glej tabelo 1).

Tabela 1: Rezultati Mann-Whitneyjevega *U* testa razlik v povprečnem številu napak pri testu Switcher

	Igralci *M* (*SD*)	Neigralci *M* (*SD*)	*U*	*z*	*p*
1. stopnja preklapljanja	0,4 (1,0)	0,5 (1,3)	2584,5	–0,018	,984
2. stopnja preklapljanja	0,5 (1,1)	0,4 (0,9)	2586,0	–0,008	,997
3. stopnja preklapljanja	0,6 (1,1)	0,4 (0,9)	2536,5	–0,262	,787

4. RAZPRAVA

Tako igralci kot igralke AVI so izkazali boljše sposobnosti mentalne rotacije od neigralcev in neigralk. Velikost učinka je bila sicer majhna (in manjša od velikosti učinka spola), vendar statistično pomembna. Sklepamo lahko, da je igranje AVI povezano z višjimi prostorskimi sposobnostmi, natančneje mentalno rotacijo, kar je skladno s predhodnimi ugotovitvami raziskav, vendar pa sta dve nedavni metaanalizi [1, 6] pokazali večje velikosti učinka. Razlog za to bi lahko iskali v tem, da smo mi raziskovali specifično sposobnost mentalne rotacije, metaanalizi pa sta vključevali splošne prostorske sposobnosti, od katerih je mentalna rotacija le del. Uporabljena različica naloge mentalne rotacije je bila časovno omejena, kar bi lahko vplivalo na rezultate. Časovni pritisk je zagotovo dodaten stresni dejavnik pri reševanju nalog in ljudje se nanj različno odzovemo. Igranje AVI pogosto poteka pod časovnim pritiskom, prav tako pa raziskave kažejo, da se igranje povezuje s hitrostjo procesiranja in krajšimi reakcijskimi časi [2], kar bi lahko vplivalo na rezultate. Zanimivo bi bilo primerjati razlike med igralci in neigralci pri časovno omejenem in neomejenem testu mentalne rotacije.

Test sledenja več objektom je mera obsega pozornosti. Predhodne raziskave [3, 8] kažejo, da z višanjem števila objektov sledenja pada povprečni delež pravilnih odgovorov, kar se je pokazalo tudi na našem vzorcu. Razlike med igralci in neigralci pri tem testu niso bile tako očitne. Statistično pomembne razlike so se pokazale po izločitvi enega udeleženca. Kaže se torej trend, da imajo igralci širši obseg pozornosti, ne moremo pa zanesljivo zaključiti, ali je razlika med skupinama večja, kot bi jo pričakovali po naključju. Takšni rezultati niso popolnoma v skladu s predhodnimi izsledki, ki kažejo, da je pozornost najbolj dosledno povezana z igranjem AVI [1, 6]. Razlog bi lahko bil v velikosti vzorca, ki je bil pri nas večji kot pri večini drugih raziskav. Prav tako bi bil lahko kriv tudi sam test – čeprav so pri enakem testu raziskovalci [3] ugotovili največje razlike med igralci in neigralci pri sledenju štirim in petim objektom, bi bilo v prihodnje dobro vključiti tudi sledenje šestim in sedmim objektom in izključiti sledenje le enemu objektu, kjer je viden učinek stropa. Tekom reševanja testa sledenja objektom se je pri nekaterih udeležencih pojavila težava, da so se njihovi dražljaji premikali prehitro v odvisnosti od hitrosti osveževanja monitorja. Zaradi tega smo morali izločiti 11 igralcev. Predpostavljamo, da so le-ti imeli zelo dobre monitorje, ravno posamezniki z dobrimi boljšimi monitorji pa so verjetno pogosti in kompetentni igralci videoiger, zato je možno, da bi bili rezultati drugačni, če bi bili vključeni tudi ti igralci.

Rezultati testa Switcher so pokazali statistično pomembne razlike med skupinama igralcev in neigralcev v reakcijskih časih pri vseh treh stopnjah preklapljanja. Velikost učinka je bila srednja do visoka, igralci so bili hitrejši tako pri predvidljivem preklapljanju med dvema in tremi pravili kot tudi pri naključnem preklapljanju. Čeprav so vse naloge reševali hitreje, pa pri tem niso naredili pomembno več napak kot neigralci. Takšni rezultati so skladni s prejšnjimi ugotovitvami o reakcijskih časih igralcev AVI [2], ki kažejo, da so igralci hitrejši, natančnost pa je v obeh skupinah primerljiva. To pomeni, da igralci na račun hitrosti ne naredijo več napak, torej ne gre za kompromis med hitrostjo in natančnostjo (angl. *speed-accuracy trade-off*). Krajši reakcijski časi igralcev kažejo na bolj razvito vidno procesiranje, pozornost in hitrejše preklapljanje med pravili ter spremembo načina reševanja, ko naloga to zahteva. Pri testu Switcher je ključna primerjava reakcijskih časov med stopnjami preklapljanja, ki poda več informacij o stroških preklapljanja kot reakcijski časi pri posameznih stopnjah preklapljanja. Igralci in neigralci se niso razlikovali v stroških preklapljanja. Število pravil (dve ali tri pravila) in (ne)predvidljivost preklopov sta na obe skupini vplivala enako. To ni skladno z drugimi korelacijskimi študijami, ki ne glede na uporabljeno paradigmo kažejo, da imajo igralci nižje stroške preklapljanja [1, 6]. Naši rezultati so bolj skladni z zaključki Karla idr. [4], ki predvidevajo, da so igralci sicer hitrejši pri preklapljanju zaradi boljšega nadzora nad selektivno pozornostjo, pri čemer pa ne gre za bolj razvite izvršilne funkcije in večjo kognitivno fleksibilnost.

Predvidevamo, da bo v prihodnosti vse manj igralcev, ki igrajo izključno AVI, hkrati pa bodo meje med različnimi zvrstmi videoiger vse manj jasne. Razvijalci v igre vključujejo priljubljene lastnosti različnih žanrov, zato ima vse več videoiger tudi nekatere lastnosti akcijskih [1]. Iz teh razlogov bi se bilo v prihodnosti bolje osredotočiti na posamezne lastnosti in kognitivne funkcije, ki jih videoigre vključujejo, in ne na specifične zvrsti. Le-te so namreč slabši indikatorji kognitivnih procesov, ki jih zahteva igranje. Na podlagi karakteristik iger, katerih igranje se povezuje z višjimi sposobnostmi, bi lahko raziskovalci tudi lažje izdelali videoigro v namene razvijanja specifičnih sposobnosti.

Za konec bi izpostavili tudi, da skupina igralcev ne vključuje izključno posameznikov, ki videoigre igrajo cele dneve oz. ki svoj čas že nekoliko nezdravo posvečajo le igranju. Zaključki, da prekomerno igranje AVI pripomore k izboljšanju kognitivnih sposobnosti, so torej napačni, in sicer iz dveh razlogov. Prvič zato, ker je bil kriterij za vključitev igranje več kot 6 ur na teden. Pri tem ne dobimo podatka o tem, kako se povezava spreminja z naraščanjem števila ur igranja. In drugič zato, ker primerjava med rednimi igralci in neigralci ne daje zaključkov o tem, ali je igranje res vzrok bolj razvitih kognitivnih sposobnosti. Morda posamezniki, ki imajo boljše kognitivne sposobnosti, igrajo več AVI, ker se v njih dobro odrežejo. Za odkrivanje, ali so AVI vzrok izboljšanja, so potrebne skrbno načrtovane in nadzorovane eksperimentalne longitudinalne študije.

5. ZAKLJUČKI

V sodobnem času veliko ljudi namenja vse več časa igranju videoiger. Posledično se kaže potreba razumeti, kako takšno početje vpliva na človeško kognicijo. Prvi korak do odgovorov je primerjava med igralci AVI in neigralci. Problem raziskave je bil ugotoviti, ali se skupini med seboj razlikujeta v obsegu pozornosti in sposobnostih mentalne rotacije in preklapljanja med nalogami. Vse to so ključne sposobnosti, ki jih uporabljamo v vsakdanjem življenju in so potrebne za uspešnost na različnih področjih. Na slovenskem vzorcu se je pokazalo, da je igranje AVI povezano s sposobnostjo mentalne rotacije in hitrostjo procesiranja vidnih informacij, medtem ko povezava z obsegom pozornosti in preklapljanjem med nalogami ni bila tako jasna. Vsekakor se kaže trend, da imajo igralci tudi ti dve sposobnosti bolje razviti, vendar so učinki majhni. Nadaljnje raziskave bi lahko razčistile vprašanja in pomanjkljivosti pričujoče študije. Poznavanje značilnosti videoiger, ki vplivajo na določene kognitivne funkcije, je ključnega pomena ne samo zato, ker so videoigre tako razširjene po celem svetu, temveč tudi zaradi potencialne uporabe v učnem in zdravstvenem kontekstu.

6. REFERENCE

[1] Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S. in Bavelier, D. 2018. Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychol. Bull.* 144, 1, 77-110. DOI = http://dx.doi.org/10.1037/bul0000130.

[2] Dye, M. W. G., Green, C. S. in Bavelier, D. 2009. Increasing Speed of Processing With Action Video Games. *Curr. Dir. Psychol. Sci.* 18, 6, 321-326. DOI = http://dx.doi.org/10.1111/j.1467-8721.2009.01660.x.

[3] Green, C. S. in Bavelier, D. 2006. Enumeration versus multiple object tracking: the case of action video game players. *Cognition.* 101, 1, 217-245. DOI = http://dx.doi.org/10.1016/j.cognition.2005.10.004.

[4] Karle, J. W., Watter, S. in Shedden, J. M. 2010. Task switching in video game players: Benefits of selective attention but not resistance to proactive interference. *Acta Psychologica.* 134, 1, 70–78. DOI = http://dx.doi.org/10.1016/j.actpsy.2009.12.007.

[5] Mueller, S. T. 2012. The PEBL Switcher Task. Pridobljeno s http://pebl.sf.net/battery.html.

[6] Sala, G., Tatlidil, K. S. in Gobet, F. 2018. Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychol. Bull.* 144, 2, 111-139. DOI = http://dx.doi.org/10.1037/bul0000139.

[7] Vandenberg, S. G. in Kuse, A. R. 1978. Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Percept. Mot. Skills.* 47, 2, 599-604. DOI = https://doi.org/10.2466/pms.1978.47.2.599.

[8] Yung, A., Cardoso-Leite, P., Dale, G., Bavelier, D. in Green, C. S. 2015. Methods to Test Visual Attention Online. *J. Vis. Exp.* 96. DOI = https://doi.org/10.3791/52470.

Regular and irregular forms: evidence from Parkinson's and Alzheimer's disease in Slovene-speaking individuals

Georgia Roumpea
Department of Comparative and General Linguistics
University of Ljubljana
georoubea@gmail.com

Maja Blesić
MEi:CogSci
Faculty of Education
University of Ljubljana
Slovenia
majablesic2@gmail.com

Dejan Georgiev
Department of Neurology, University Medical Centre, Ljubljana
Slovenia
georgievdejan@gmail.com

Christina Manouilidou
Department of Comparative and General Linguistics
University of Ljubljana
Slovenia
Christina.Manouilidou@ff.uni-lj.si

ABSTRACT

According to the Declarative/Procedural Model, declarative and procedural memory play a specific role in the production of irregular and regular forms (REG: *talk-ed*, IRR: *went*, respectively). In Parkinson's disease where procedural memory is impaired, and Alzheimer's disease, where declarative memory limitations are manifested, the production of (ir)regular forms has been widely investigated leading to contradictory results. The current study reports evidence from Slovene-speaking PD and AD patients, by examining the production of (ir)regular forms in the formation of number (singular vs. plural), tense (present, past, future) and grammatical aspect (perfective vs. imperfective). Participants performed worse than the control group, but no dissociation between regular and irregular forms was observed, suggesting that declarative and procedural memory are possibly involved in linguistic process, but they might not play a crucial role in the production of (ir)regular forms.

Keywords

Parkinson's disease, Alzheimer's disease, (ir)regular morphology, Declarative/Procedural Model.

1. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease characterized by impairment in temporal lobe structures [1]. Dysfunction in declarative memory (semantic and episodic), rooted in temporal lobe structures, is manifested early on the disease. Procedural memory (rooted in basal ganglia) is considered to be relatively preserved [2]. Language abilities are affected during all stages of the disease with patients having difficulties in both production and comprehension of grammatical and semantic aspects of language. Fyndanis et al. [3] report impaired tense and grammatical aspect (perfective "*played*", imperfective "I was playing") production and comprehension in Greek-speaking mild-AD patients. Roumpea et al. [4] observed similar language performance in mild-AD patients. Concerning semantic aspects of language, Kim and Thompson [5] report noun and verb naming deficits in AD. Language impairment in individuals with AD results from declarative memory (mainly semantic memory) limitations [1] and working memory impairment [6].

Parkinson's disease (PD) is a neurodegenerative disorder characterized by loss of dopamine in the basal ganglia and degeneration of subcortical frontal structures. These areas sustain procedural memory which has been found to be impaired in PD [2]. Declarative memory (temporal lobe) is considered to be preserved [2]. Macoir et al. [7] mention that individuals with PD mainly display motor system dysfunction, but language deficits are also observed (e.g. difficulties in sentence comprehension and production), while semantic features (e.g. word recognizing) remain unimpaired. Basal ganglia impairment has been assumed to affect PD patients' language abilities. More specifically, PD patients' language limitations are attributed to degraded procedural memory which is responsible for the computation of rule-based linguistic procedures.

2. LINGUISTIC BACKGROUND AND BACKGROUND RESEARCH

2.1. Regular and Irregular morphology in Slovene

Slovene is a language with rich morphology and manifests both regular and irregular forms in multiple domains, such as number (singular vs. plural), tense (present tense) and grammatical aspect (perfective vs. imperfective). Grammatical aspect conveys information about how a situation took place in time. Perfective aspect (*I walked*) presents a non-durative situation, while imperfective presents (*I was walking*) a durative situation.

The regular formation of the above categories is either by suffixation or by prefixation [8]. Suffixation is a morphological operation where a morpheme (e.g. –ed) is attached to the end of a word (stem e.g. walk) [walk + ed → walked (English past tense)]. Prefixation is a morphological operation where a morpheme (e.g. un-) is attached to the front part of a word (stem e.g. lock) (un + lock → unlock). While present tense and number are regularly formed by suffixation (delati$_{inf}$ – delam$_{1sing}$ "to work – I work",

$miza_{sing}$ – $miz\mathbf{e}_{pl}$ "table – tables"), in the formation of *aspect*, the corresponding perfective form of an imperfective infinitive is usually formed by prefixation ($risati_{imperf}$ – $\mathbf{na}risati_{perf}$ "to draw – to finish drawing"). However, irregular forms in the formation of tense, number and grammatical aspect are also observed, such as as "iti_{inf} – $grem_{1sing}$ "to go – I go", $človek_{sing}$ – $ljudje_{pl}$ "human - humans" and $metati_{imperf}$ – $vreči_{perf}$ "to throw – to finish throwing", respectively.

2.2. Background research and predictions for PD and mAD

Concerning the processes of regular and irregular forms, one of the proposed models is the Declarative/Procedural Model (D/PM) by Ullman et al. [9]. According to D/PM, the declarative memory (temporal lobe structures) stores and processes lexical information and is responsible for the production of irregular forms (go → **went**), while procedural memory (basal ganglia) processes grammatical rules, thus it is responsible for the production of regular forms (*walk → walk**ed***).

The process of regular and irregular forms, as suggested by the D/PM, has been widely investigated with studies leading to mixed results. Ullman et al. [9] found irregular forms of English past tense (*I taught*) to be impaired in AD individuals, but better preserved in PD, while regular forms (*I played*) were better preserved in AD individuals compared to PD. The authors claimed that degeneration of declarative memory in AD and impaired procedural memory in PD might explain these results (D/PM). The same findings and rationale are reported for PD and AD patients in Cameli et al. [10].

However, there are several studies which failed to replicate the D/PM. Macoir et al. [7] found that French-speaking PD patients' performance did not differ for regular and irregular verbs in experimental conjugation tasks. The authors suggest that basal ganglia, where procedural memory is rooted, interfere with language processing but do not play a specific role in verb production as proposed by the D/PM. Similarly, Terzi et al. [11] and Penke and Wimmer [12] report that Greek-speaking and German-speaking PD individuals showed no dissociation between regular and irregular verbs. The authors claimed that there was no evidence for a selective deficit affecting the production of regular forms, suggesting that basal ganglia and procedural memory do not play a crucial role in the production of regular forms as proposed by the D/PM.

Motivated by the above contradictory results, in the present study we investigate the production of regular and irregular forms in Slovene language in the categories of number, tense and grammatical aspect. While the production of regular and irregular forms has been examined widely in other languages (e.g. English), to our knowledge there is no evidence from Slavic languages. This study is one of the first attempts to investigate the issue of nominal and verbal (ir)regularity in Slovene. Furthermore, we will examine whether the production of Slovene regular and irregular morphology is supported by the D/PM [9].

Concerning AD, we expect that declarative memory decline will lead participants to have difficulties in producing the irregular forms of all the under examination categories (number, tense and grammatical aspect). The irregular forms are supposed to be retrieved directly from the declarative memory as they are not subject to grammatical rules. On the other hand, procedural memory impairment in PD participants might lead them to perform better in producing irregular forms of number, tense and grammatical aspect compared to regular ones. This is expected based in the fact that in order to produce regular forms application of grammatical rules is needed. Procedural memory limitations might cause difficulties in applying grammatical rules to PD. Finally, differences among the categories of number, tense and grammatical aspect might arise due to the different morphological operations (suffixation or prefixation) used in their formation.

3. METHODOLOGY

3.1 Participants

Five individuals with no neurological impairments (4 females, 1 male), 5 no-dementia PD (all males) and 6 mild-AD (henceforth mAD, all females) all native-Slovene speakers participated in this study. PD and mAD participants were recruited at the Neurological clinic of Ljubljana, all diagnosed by a qualified neurologist, while the healthy participants were recruited from an Elderly Care House in Ljubljana "Dom starejših občanov Ljubljana-Bežigrad". Participants were matched when it comes to age and years of education. The Mini Mental State Examination (MMSE) was administered to all participants to collect more information about their cognitive profile. Table 1 provides detailed information on participants' demographics as well as their scores in the neuropsychological task.

Table 1: Participants' demographic and neuropsychological information. Standard deviations are given in parentheses. T-scores, p-values and Degrees of Freedom (df) from independent samples t-tests comparing the groups are also reported.

	PD	**mAD**	**Control Group**
Mean age	77.6	82.5	73.6
Education level	13.2 (1.7)	12.8 ()	12.8 (4.3)
MMSE	27.0 (1.8)	19.5 (3.0)	28.2 (1.3)
	Statistical Comparisons		
	PD vs. mAD	**PD vs. Control group**	**mAD vs. Control Group**
Mean age	t= .943, p= .370, df= 9	t= -.618, p= .554, df= 8	t= 1.833, p= .100, df= 9
Education level	t= -.619, p= .551, df= 9	t= .189, p= .855, df= 8	t= -.218, p= .832,df= 9
MMSE	t= -4.817, p< .01, df= 9	t= 1.177, p= .273, df= 8	t= -5.960 p< .01, df= 9

3.2 Stimuli and Experimental task

Our stimuli consist of 23 verbs [11 regulars, 12 irregulars and 6 nouns (3 regulars, 3 irregulars)]. A sentence-completion task was designed and it included 29 pairs of source sentences (SS) and

target sentences (TS): 6 of them designed to test number, i.e. singular vs. plural (3 regulars, 3 irregulars), 14 tested present tense (6 regulars, 8 irregulars) and 9 tested aspect (5 regulars, 4 irregulars/ 5 perfective, 4 imperfective). The SS and the TS were presented simultaneously to the participants. The SS differed from the TS only to the point that it was necessary to trigger the production of the target verb or noun forms. The sentence in (1) is an example of producing regular tense.

(1) SS: Hoditi v šolo je pomembno. (To go to school is important)
TS: Maja zdaj hodi v prvi razred. (Now, Maja is now going to the first grade).

3.3 Procedure

Power Point was used to present the experimental materials to the participants. Each pair of sentences (SS and TS) was presented separately to the participants. At the beginning of the experimental procedure, participants were provided with instructions of how to complete the task. 3 pairs of sentences that were not included in the stimuli were used as examples in order to get participants familiar with the task. Participants' responses during the trial period were not taken into account in the analysis. Participants had to complete the right form of the missing noun or verb. The task was off-line and participants had as much time as they needed in order to complete the sentence.

4. RESULTS

For the statistical analysis of the results we performed the Fisher's exact test, a non-parametric statistical test for small samples that are not normally distributed. In all statistical comparisons, participants' responses (correct and incorrect responses) were treated as the dependent variable, while the different participants' groups (PD, mAD, controls), number (singular vs. plural), grammatical aspect (perfective, imperfective) and tense (present) were treated as the independent variables.

Groups' percentages of correct responses are illustrated in Figure 1. mAD group performed lower (74% correct responses) than PD group (94.5% correct responses) and control group (98.5% correct responses), with mAD being statistically worse ($p < .01$, in all comparisons) compared to both PD and control group, while PD group performed equally well ($p = .103$).

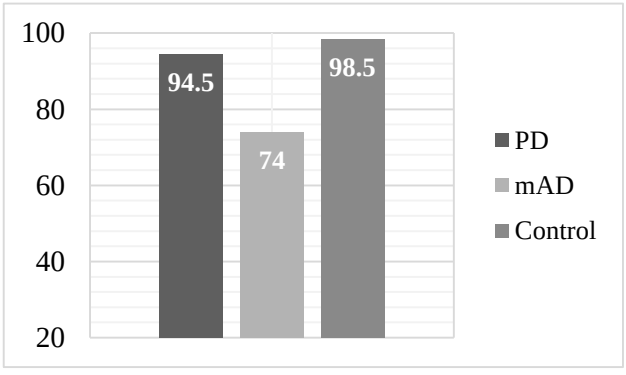


Figure 1: Performance (% correct) in the sentence-completion task of PD, mAD and control groups.

Participants' correct responses in regular and irregular forms are presented in Figure 2. No dissociation between regular and irregular verbs was observed both for individuals with mAD and PD ($p > .05$ and $p = .863$, respectively).

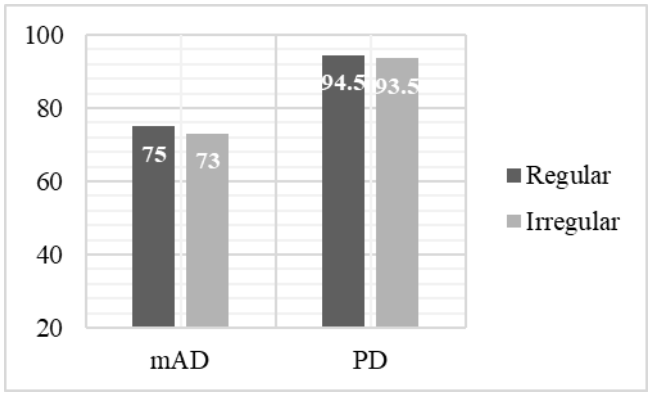


Figure 2: Individuals' with PD and mAD performance (% correct) in regular and irregular forms.

mAD performance (% correct) in regular and irregular forms with respect to number, tense and grammatical aspect is illustrated in Figure 3. No statistically significant difference between regular and irregular forms was found for number, tense and grammatical aspect ($p > .05$, in all comparisons). Concerning the performance in the different grammatical categories, the highest score was achieved for number, followed by tense and aspect, where participants performed lower. This difference reached significance, with aspect being statistically worse compared to both tense and number ($p = .039$, in all comparisons).

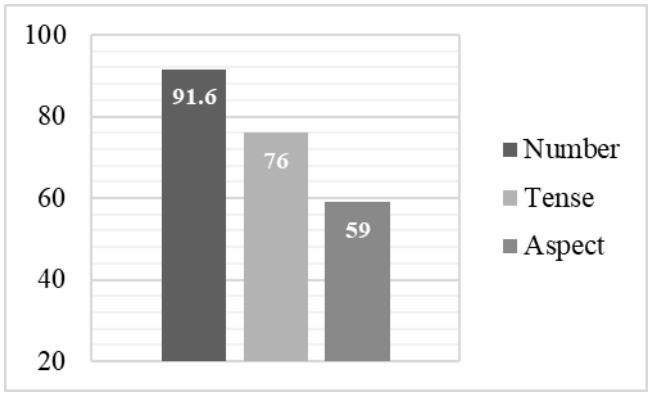


Figure 3: Individuals' with mAD performance (% correct) with respect to plural, tense and aspect.

Error analysis revealed that mAD participants had no difficulty in producing the present tense of regular (*delati*$_{inf}$ – *dela**m***$_{1sing}$) and irregular verbs (iti$_{inf}$ – grem$_{1sing}$). In detail, participants were able to use correctly the appropriate suffix to form the present tense of both regular and irregular verbs. The most frequent mistake in the category of tense was the substitution of the target irregular form with a regular one of a verb that was semantically close to the target one and correctly formed [*hodim* (regular, I am walking) -> *grem* (irregular, I am going)]. Regarding aspect, in regular verbs mAD individuals tended to produce the imperfective form of the verb instead of the perfective target (**na**pisala$_{perfective}$ → pisala$_{imperfective}$). In irregular verbs participants tended either to produce the no-target aspect category (jemali$_{imperfective}$ →

vzeli$_{perfective}$) or to substitute the target verb with another irregular semantically related verb.

5. DISCUSSION

In this study we investigated the production of regular and irregular forms as they are manifested in the categories of number (singular vs. plural), tense (present tense) and grammatical aspect (perfective and imperfective). With this study, we attempt to contribute to the existing literature on the irregularity issue by bringing evidence from Slovene, where (ir)regularity is manifested in multiple domains.

Concerning the overall performance of mAD, PD and control groups, only individuals with mAD were found statistically impaired in the production of regular and irregular forms compared to both PD and control groups, while PD's group performance was equal to the control's group. Moreover, in mAD group no dissociation between regular and irregular forms was observed. The current findings do not support the D/PM, contra to studies that replicated it [10]. On the other hand, our results are in line with studies that failed to support the D/PM and suggested that declarative and procedural memory are involved in language processing but they might not play a specific role in (ir)regular forms production [7].

The lack of dissociation between regular and irregular forms in Slovene might be explained by the fact that Slovene is a morphologically rich language. In detail, contrary to English, grammatical rules are applied both to regular and irregular forms. Suffixation is applied to form the regular plural (miza$_{sing}$ – miz**e**$_{pl}$ "table – tables") and present tense (delati$_{inf}$ – dela**m**$_{1sing}$ "to work – I work"), while prefixation to form the perfective aspect (risati$_{imperf}$ – **na**risati$_{perf}$ "to draw – to finish drawing"). However, suffixation is also applied to irregular forms of number, present tense and grammatical aspect due to the fact that Slovene manifests agreement (case, person, gender etc.), thus after retrieving the irregular forms from declarative memory, speakers need also to apply grammatical rules according to agreement (see Terzi et al [11] for a similar explanation for Greek-speaking PD patients).

Regarding individuals with mAD performance in number, tense and grammatical aspect, difficulties in completing the perfective form of regular verbs, might suggest impairment in producing the aspectual prefix. Morphology of number and tense (both formed by suffixation) seems to be spared. These findings are in line with Kavé and Levy [13] who report impaired prefixation and preserved suffixation in AD. Nonetheless, aspect has been found to be impaired compared to tense, in languages that use suffixation to form it (Fyndanis et al. [3] for Greek). Thus, further research is needed to clarify whether morphological or other factors (e.g. semantics) might interfere with clinical populations' ability to produce aspect.

To sum up, the current study is one of the first attempts to investigate the issue of irregularity in Slovene. Our findings do not support the proposed D/PM [9] for the production of regular and irregular forms, suggesting that PD and mAD individuals' difficulties in regular or irregular forms are not directly connected with declarative and procedural memory limitations. Finally, due to the small sample of participants and stimuli further research is needed to come up with more accurate results regarding the issue of irregularity in Slovene.

6. REFERENCES

[1] Braaten, A. J., Parsons, T. D., McCUE, R., Sellers, A., and Burns, W. J. 2006. Neurocognitive Differential Diagnosis of Dementing Diseases: Alzheimer's Dementia, Vascular Dementia, Frontotemporal Dementia, And Major Depressive Disorder. *International Journal of Neuroscience* 116, 11 (2006), 1271–1293. DOI= http://dx.doi.org/10.1080/00207450600920928

[2] Gabrieli, J. D., Corkin, S., Mickel, S. F., and Growdon, J. H. 1993. Intact acquisition and long-term retention of mirror-tracing skill in Alzheimer's disease and in global amnesia. *Behavioral Neuroscience* 107, 6 (1993), 899–910. DOI= http://dx.doi.org/10.1037//0735-7044.107.6.899

[3] Fyndanis, V., Manouilidou, C., Koufou, E., Karampekios, S., and Tsapakis, E. M. 2013. Agrammatic patterns in Alzheimer's disease: Evidence from tense, agreement, and aspect. *Aphasiology* 27, 2 (2013), 178–200. DOI= http://dx.doi.org/10.1080/02687038.2012.705814

[4] Roumpea, G., Nousia, A., Stavrakaki, S., Nasios, G., and Manouilidou, C. 2019. Lexical and grammatical aspect in Mild Cognitive Impairment and Alzheimer's disease. Selected papers on theoretical and applied linguistics, 23, (2019), 381-397. DOI= http://ejournals.lib.auth.gr/thal/article/view/7355.

[5] Kim, M., and Thompson, C. K. 2004. Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and Language* 88, 1 (2004), 1–20. DOI= http://dx.doi.org/10.1016/s0093-934x(03)00147-0

[6] Kensinger, E. A., Shearer, D. K., Locascio, J. J., Growdon, J. H., and Corkin, S. 2003. Working memory in mild Alzheimer's disease and early Parkinson's disease. *Neuropsychology*, 17(2), (2003), 230.

[7] Macoir, J., Fossard, M., Mérette, C., Langlois, M., Chantal, S., and Auclair-Ouellet, N. 2013. The role of basal ganglia in language production: evidence from Parkinson's disease. *Journal of Parkinson's disease*, 3(3), (2013), 393-397. DOI= http://doc.rero.ch/record/309035/files/Fossard_Marion_The_Role_of_Basal_Ganglia_in_Language_Production_20180418.pdf

[8] Herrity, P. 2016. Slovene: a comprehensive grammar, London: Routledge.

[9] Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., and Pinker, S. 1997. A Neural Dissociation within Language: Evidence that the Mental Dictionary Is Part of Declarative Memory, and that Grammatical Rules Are Processed by the Procedural System. *Journal of Cognitive Neuroscience* 9, 2 (1997), 266–276. DOI= http://dx.doi.org/10.1162/jocn.1997.9.2.266

[10] Cameli, L., Phillips, N. A., Kousaie, S., and Panisset, M. 2005. Memory and language in bilingual Alzheimer and Parkinson patients: Insights from verb inflection. In ISB4: *Proceedings of the 4th International Symposium on Bilingualism*, (2005), (pp. 452-476). DOI=

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.653.9612andrep=rep1andtype=pdf

[11] Terzi, A., Papapetropoulos, S., and Kouvelas, E. D. 2005. Past tense formation and comprehension of passive sentences in Parkinson's disease: Evidence from Greek. *Brain and Language* 94, 3 (2005), 297–303. DOI= http://dx.doi.org/10.1016/j.bandl.2005.01.005

[12] Penke, M., and Krause, M. 2004. Regular and irregular inflectional morphology in German Williams syndrome. *Williams Syndrome across Languages Language Acquisition and Language Disorders* (2004), 245–270. DOI= http://dx.doi.org/10.1075/lald.36.15pen

[13] Kavé, G., and Levy, Y. 2004. Preserved Morphological Decomposition in Persons with Alzheimer's Disease. *Journal of Speech, Language, and Hearing Research* 47, 4 (2004), 835–847. DOI= http://dx.doi.org/10.1044/1092-4388(2004/062)

Dva pristopa k opredelitvi in preučevanju delovnega spomina

Two approaches to defining and studying working memory

Anka Slana Ozimič
Mind and Brain Lab, Department of Psychology, Faculty of Arts, University of Ljubljana
Aškerčeva 2
1000 Ljubljana
anka.slana@ff.uni-lj.si

POVZETEK

Prispevek se osredotoča na vidno-prostorski delovni spomin in predstavlja dva odmevnejša teoretska okvirja za preučevanje delovnega spomina: Baddeleyjev in Hitchev multikomponentni model delovnega spomina in Cowanov model vpetih procesov. Opisuje, kako modela razumeta in opredeljujeta delovni spomin, njegove komponente in vključene procese ter kakšna je vloga le-teh pri omejevanju njegove kapacitete. Na koncu predstavimo, kako se komponente modelov povezujejo z možganskimi sistemi.

Ključne besede

Delovni spomin, multikomponenti model, model vpetih procesov, reprezentacije, aktivno vzdrževanje.

ABSTRACT

The paper focuses on visual-spatial working memory and presents two prominent theoretical frameworks for the study of working memory: Baddeley's and Hitch's multicomponent model of working memory and Cowan's model of embedded processes. It describes how models understand and define working memory, its components, and the processes involved, and their role in limiting its capacity. In the end we present how the components of the models relate to the brain systems.

Keywords

Working memory, multicomponent model, model of embedded processes, representations, active maintenance.

1. UVOD

Delovni spomin je sposobnost vzdrževanja in aktivnega manipuliranja z informacijami potrebnih za dosego trenutnega cilja, medtem ko se pojem kratkoročnega spomina nanaša na enostavno začasno shranjevanje informacij [1]. Delovni spomin je ena izmed temeljnih kognitivnih sposobnosti, ki omogoča opravljanje vsakodnevnih aktivnosti. Visoko korelira s splošno inteligentnostjo [2] in je pogosto oškodovan pri boleznih možganov [3], njegov upad pa je značilen tudi za zdravo staranje [4]. Zaradi njegove osrednje vloge v kogniciji je raziskovanje temeljnih mehanizmov delovnega spomina izrednega pomena za razumevanje človeške kognicije.

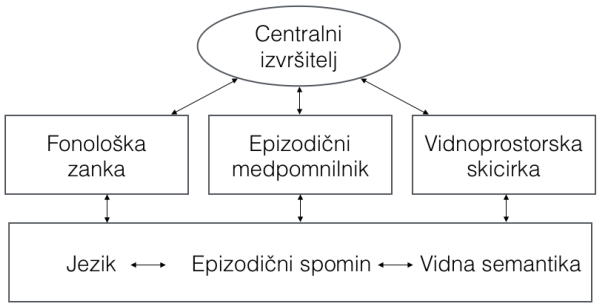
V preteklosti se je glavnina raziskav osredotočala na verbalni delovni spomin [5], vrsto delovnega spomina, v katerem informacije hranimo v obliki besed in zvoka. To pa ni edina oblika informacij, ki jih je potrebno hraniti pri izvedbi vsakodnevnih opravil. Pogosto se zanašamo na vidne ali prostorske informacije, ki nam omogočjo vzdrževanje podob in položajev v okolju, ko le-ti niso neposredno dostopni v našem vidnem polju.

Eno izmed ključnih raziskovalnih vprašanj pri preučevanju vidno-prostorskega delovnega spomina se nanaša na mehanizme, ki so temelj njegovi omejeni kapaciteti. Raziskave kažejo, da je v vidnem delovnem spominu v danem trenutku možno vzdrževati 3–4 enote informacij [6]. V sklopu preučevanja kapacitete so aktualne številne znanstvene diskusije, npr. ali so enote omejene kapacitete vidnega delovnega spomina reprezentacije integriranih objektov ali individualnih lastnosti [7], ali omejitve izhajajo iz modalno-specifičnih shramb ali iz omejitev v procesih pozornosti [6].

Kaj omejuje kapaciteto delovnega spomina, skušajo pojasniti številni modeli, ki so se razvili v okviru kognitivne psihologije in nevroznanstvenega preučevanja in razlagajo njegovo strukturo, vključene komponente in procese. Dva izmed odmevnejših teoretskih okvirjev za preučevanje delovnega spomina sta Baddeleyjev in Hitchev [1, 8] multikomponentni model delovnega spomina in Cowanov model vpetih procesov [2].

2. MULTIKOMPONENTNI MODEL DELOVNEGA SPOMINA

Baddeley in Hitch [8] v svojem modelu delovni spomin opisujeta kot hipotetični sistem omejene kapacitete, ki omogoča začasno shrambo in manipulacijo informacij, potrebnih za izvedbo številnih kognitivnih aktivnosti. Model vključuje več komponent: "suženjske" komponente za začasno shranjevanje informacij, ki poleg shramb vključujejo procese za osveževanje informacij, in izvršilno komponento, ki aktivno upravlja z informacijami v delovnem spominu ter preko nadzora procesov pozornosti opredeljuje vnos in iznos iz shramb (Slika 1).



Slika 1: Multikomponentni model delovnega spomina. V model so vključene komponente za kratkoročno shranjevanje informacij (fonološka zanka, epizodični medpomnilnik in vidnoprostorska skicirka), ki se povezujejo z vsebinami iz dolgoročnega spomina, ter centralni izvršitelj, ki le-te nadzira in določa, katere informacije vanje vstopajo.

Sprva sta bila v model vključena dva sistema za shranjevanje informacij [8]: fonološka zanka, zadolžena za vzdrževanje informacij v fonološki obliki, in vidno-prostorska skicirka, ki hrani vidne in prostorske informacije. Baddeley in Hitch sta nadalje predvidela, da fonološka zanka sestoji iz pasivne shrambe omejene kapacitete (fonološka shramba) in aktivnega procesa za osveževanje informacij (artikulatorni kontrolni proces), ki ponovno aktivira in pomaga preprečevati propadanje spominskih sledi. Logie [9] je s preučevanjem vidno-prostorskega delovnega spomina nadgradil Baddeleyev in Hitchev model. Analogno fonološki zanki je za vidno-prostorsko skicirko predvidel, da tudi ta sestoji iz pasivne shrambe za vidne informacije (vidna shramba) in aktivnega sistema za osveževanje in prostorsko manipulacijo informacij (notranja skicirka) [9].

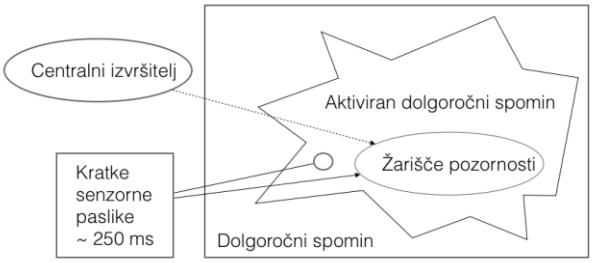
Poleg fonološke zanke in vidno-prostorske skicirke sta Baddeley in Hitch [8] v izvirni model vključila centralnega izvršitelja, ki igra vlogo nadzornika in upravlja celotni sistem ter omogoča manipulacijo z informacijami, hranjenimi v shrambah. Njegova glavna naloga je nadzor pozornosti—usmerjanje, razdeljevanje in preklapljanje pozornosti. Medtem ko imajo sistemi za shranjevanje omejeno spominsko kapaciteto, je kapaciteta centralnega izvršitelja omejena s pozornostnimi viri [1].

Četrta komponenta, epizodični medpomnilnik, je bil modelu dodan kasneje [10]. Njegova naloga je hranjenje integriranih informacij različnih modalnosti v obliki kratkih epizod in povezovanje z dolgoročnim spominom [10]. Začetna predpostavka je bila, da je shranjevanje integriranih informacij močno odvisno od pozornostnega nadzora centralnega izvršitelja, vendar so empirična spoznanja potrdila, da je epizodični medpomnilnik — v nasprotju s fonološko zanko in vidno-prostorsko skicirko, ki vsebujeta lastne mehanizme osveževanja informacij — pasivna struktura, ki ne potrebuje pozornosti centralnega izvršitelja za vzdrževanje informacij, temveč da je le-ta potrebna za obrambo pred motečimi dražljaji [1].

3. MODEL VPETIH PROCESOV

Ideja modelov stanj [11], katerih najbolj vidni predstavnik je Cowanov model vpetih procesov [2], je podati bolj splošen opis sistema delovnega spomina v okviru procesov, ki jih vključuje. Model vpetih procesov namesto sistemov za začasno shrambo in centralnih izvršilnih procesov, ki le-te nadzirajo, delovni spomin vidi kot sistem za nadzor usmerjanja pozornosti na trenutno aktivirane vsebine epizodičnega in semantičnega dolgoročnega spomina. Znotraj te perspektive je usmeritev pozornosti k notranjim reprezentacijam, ki so shranjene bodisi v dolgoročnem spominu [npr. 2, 12] bodisi vzpostavljene preko senzoričnih in motoričnih sistemov [npr. 13, 14], podlaga kratkoročnemu vzdrževanju informacij v delovnem spominu

Model vpetih procesov predpostavlja, da je delovni spomin aktivni del dolgoročnega spomina in pri tem opredeljuje dve temeljni komponenti [2]: aktiviran dolgoročni spomin in žarišče pozornosti (Slika 2). Aktiviran dolgoročni spomin predstavlja zbirko reprezentacij, ki so za omejen čas v posebej dostopnem stanju. Nima omejene kapacitete v smislu možnega števila sočasno aktiviranih reprezentacij, temveč je omejen s časom in interferenco med reprezentacijami. Druga komponenta je žarišče pozornosti, ki predstavlja podmnožico reprezentacij v aktiviranem dolgoročnem spominu. Žarišče pozornosti je na informacije usmerjeno bodisi avtomatično z orientacijskim refleksom na podlagi sprememb v okolju bodisi voljno s pomočjo centralnih izvršilnih procesov. Medtem ko so senzorne reprezentacije lahko aktivirane avtomatično, je za integracijo reprezentacij in nove povezave v delovnem spominu potrebna pozornost.



Slika 2: Model vpetih procesov. Predstavljena informacija najprej sproži kratko senzorno pasliko. Ta aktivira relevantne reprezentacije v dolgoročnem spominu (senzorne in kategorične). Nekatere od teh informacij preidejo v žarišče pozornosti bodisi zaradi avtomatičnih odzivov bodisi s pomočjo centralnih izvršilnih procesov, naravnanimi v skladu s cilji tekoče naloge.

4. PODOBNOSTI IN RAZLIKE MED MODELOMA

Čeprav se modela v izhodiščih razlikujeta—multikomponentni model deli komponente glede na vsebino shranjenih informacij (jezikovne, vidno-prostorske, integrirane informacije) in vključenih procesov (hramba informacij, osveževanje informacij, aktivna manipulacija s pomočjo izvršilnih procesov), medtem ko se modeli stanj namesto na obliko informacij osredotočajo na funkcijo — sta v svojem bistvu komplementarna [9]. Oba predvidevata dva sistema, ki sta vključena v vzdrževanje informacij. En omogoča vzpostavitev in hrambo reprezentacij informacij, medtem ko drugi njihovo aktivno vzdrževanje. Multikomponentni model predvideva [1], da so reprezentacije vzpostavljene in vzdrževane v spominskih shrambah "suženjskih" komponent za shranjevanje informacij (fonološka shramba, vidna shramba) in da je njihovo aktivno vzdrževanje omogočeno preko procesov osveževanja (npr. artikulatorni kontrolni proces, notranja skicirka) ter s pomočjo centralnih izvršilnih procesov, ki aktivno obdelujejo informacije v teh shrambah [15]. V modelu vpetih procesov [2] kot v modelih stanj v splošnem [11] so reprezentacije vzpostavljene bodisi znotraj sistemov za dolgoročni spomin bodisi znotraj senzoričnih in motoričnih sistemov, medtem ko centralni izvršilni procesi omogočajo njihovo aktivno vzdrževanje v žarišču pozornosti.

Oba modela torej kažeta na pomembnost obeh vključenih sistemov, razlikujeta pa se po tem, kako razumeta vlogo obeh sistemov pri omejevanju kapacitete delovnega spomina. Multikomponentni model delovnega spomina omejeno kapaciteto razume kot emergentni pojav delovanja več komponent [9]. Čeprav centralni izvršitelj nima omejene spominske kapacitete za hranjenje informacij, je omejen s pozornostnimi viri. Sistemi za shranjevanje informacij so na drugi strani omejeni s tem, koliko informacij lahko v njih shranimo. Tako za fonološko zanko kot vidno-prostorsko skicirko je značilen propad reprezentacij s časom (približno dve sekundi), če te niso ustrezno osvežene. Model predpostavlja, da ima fonološka shramba omejeno spominsko kapaciteto [2], proces osveževanja pa nima omejitve v smislu števila enot, ki jih lahko artikulira, temveč lahko vsebine osvežuje, dokler so te dostopne v shrambi. Analogno fonološki zanki, Logie [9] za vidno-prostorsko skicirko predpostavlja, da je vsebina vidne shrambe omejena z vidno kompleksnostjo reprezentacij (številom dražljajev v naboru, številom celic v vidni matriki), ki propadejo, če vsebine niso osvežene s pomočjo notranje skicirke, katere kapaciteta je omejena z dolžino niza informacij (npr. položajev), ki ga beleži [9].

Modeli stanj [2, 11] na drugi strani predvidevajo, da omejitve kapacitete delovnega spomina primarno izhajajo iz omejene

kapacitete žarišča pozornosti. Centralni izvršitelj omogoča aktivno vzdrževanje v žarišču pozornosti samo za omejeno število reprezentacij v aktiviranem dolgoročnem spominu. Čeprav ima le-ta neomejeno spominsko kapaciteto, je aktivacija relevantnih reprezentacij v aktiviranem dolgoročnem spominu omejena s časom in z interferenco [2].

5. MODELA V POVEZAVI Z MOŽGANSKIMI SISTEMI

Čeprav sta predstavljena modela mišljena kot konceptualni opis strukture in procesov delovnega spomina in njun namen ni preslikava komponent na možganske sisteme, sta skladna s spoznanji nevrofizioloških raziskav, ki kažejo, da imajo posteriorna in prefrontalna področja možganske skorje različno vlogo pri kratkoročnem vzdrževanju vidno-prostorskih informacij [16]. Študije kažejo, da so posteriorna področja možganske skorje tista, ki so vključena pri vzpostavljanju in/ali kratkoročnem shranjevanju vidnih oz. prostorskih reprezentacij [17], medtem ko prefrontalne regije nadzirajo usmerjanje pozornosti za njihovo aktivno vzdrževanje [18]. Vloga prefrotnalnih področij se tako sklada s Cowanovimi izvršilnimi procesi, ki nadzirajo, kaj je v žarišču pozornosti, in Baddeleyevimi sistemi osveževanja, ki ohranjanjo in reciklirajo vsebino iz shramb, ter centralnim izvršiteljem, ki skrbi za nadzor vsebin v shrambah. Posteriorne regije se v okviru Cowanovega modela povezujejo tako z aktiviranim dolgoročnim spomin kot tudi žariščem pozornosti [2], medtem ko se v okviru multikomponentnega modela posteriorne regije povezujejo z vsebino v komponentno-specifičnih shrambah, vključno z epizodičnim medpomnilnikom [1].

Model vpetih procesov predpostavlja, da v delovnem spominu vzdržujemo aktivirane vsebine iz dolgoročnega spomina in vsebine, vzpostavljene preko senzoričnih in motoričnih sistemov, medtem ko multikomponentni model predvideva, da je senzorika ločena od reprezentacij v shrambah. Novejše študije kažejo [za pregled glej 15], da področja za shranjevanje v posteriornih regijah niso edinstvena delovnemu spominu, temveč temeljijo na istih mehanizmih, ki so vključeni v reprezentacije informacij v zaznavi [19, 20], kar se sklada z modelom vpetih procesov in ugotovitvijo, da tako zaznava kot kratkoročno ter dolgoročno shranjevanje informacij temelji na delovanju istih anatomskih regij [19].

6. ZAKLJUČEK

Ideja modelov delovnega spomina je torej hipotetični prikaz njegove strukture in delovanja. Čeprav modela uspešno pojasnita mnoge kognitivne pojave, povezane s kratkoročnim shranjevanjem informacij in njihovo aktivno manipulacijo ter se v mnogih vidikih smiselno povezujeta z delovanjem možganov, njun namen ni pojasniti vseh njegovih vidikov. Razumemo ju lahko kot delovno platformo, ki jo je v skladu z empiričnimi spoznanji potrebno razvijati naprej.

7. OPOMBA AVTORJEV

Prispevek je nastal v okviru raziskovalnega projekta J3-9264 in raziskovalnega programa P5-0110, ki ga je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

8. LITERATURA

[1] Baddeley, A. (2012). Working memory: theories, models, and controversies. Annual Review of Psychology, 63, 1–29. DOI= http://doi.org/10.1146/annurev-psych-120710-100422

[2] Cowan, N. (2005). Working Memory Capacity. Hove, East Sussex, UK: Psychology Press.

[3] Goldman-Rakic, P. S. (1994). Working memory dysfunction in schizophrenia. The Journal of neuropsychiatry and clinical neurosciences, 6(4):348–357.

[4] Park, D. C. & Festini, S. B. (2017). Theories of Memory and Aging: A Look at the Past and a Glimpse of the Future. The Journals of Gerontology Series B, Psychological Sciences and Social Sciences, 72(1), 82–90. DOI= http://doi.org/10.1093/geronb/gbw066

[5] Repovš, G. & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. Neuroscience, 139(1), 5–21. DOI= http://doi.org/10.1016/j.neuroscience.2005.12.061

[6] Cowan, N. (2010). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? Current Directions in Psychological Science, 19(1), 51–57. DOI= http://doi.org/10.1177/0963721409359277

[7] Schneegans, S. & Bays, P. M. (2019). New perspectives on binding in visual working memory. British Journal of Psychology, 110(2), 207-244. DOI= https://doi.org/10.1111/bjop.12345

[8] Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), Recent advances in learning and motivation, Vol.8 (pp. 47–90). New York: Academic Press.

[9] Logie, R. H. (2011). The Functional Organization and Capacity Limits of Working Memory. Current Directions in Psychological Science, 20(4), 240–245. DOI= http://doi.org/10.1177/0963721411415340

[10] Baddeley, A. (2000). The episodic buffer: a new component of working memory? Trends in Cognitive Sciences, 4(11), 417–423.

[11] D'Esposito, M. & Postle, B. R. (2015). The cognitive neuroscience of working memory. Annual Review of Psychology, 66, 115–142. DOI= http://doi.org/10.1146/annurev-psych-010814-015031

[12] Oberauer, K. (2009). Design for a Working Memory. In The Psychology of Learning and Motivation (Vol.51, pp. 45 – 100). Elsevier.

[13] Magnussen, S. (2000). Low-level memory processes in vision. Trends in Neurosciences, 23(6), 247–251. DOI= http://doi.org/10.1016/s0166-2236(00)01569-1

[14] Zaksas, D., Bisley, J. W. in Pasternak, T. (2001). Motion information is spatially localized in a visual working-memory task. Journal of Neurophysiology, 86(2), 912–921. DOI= http://doi.org/10.1152/jn.2001.86.2.912

[15] Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A. & Jonides, J. (2012). A meta-analysis of executive components of working memory. Cerebral cortex, 23(2), 264-282. DOI= http://doi.org/10.1093/cercor/bhs007

[16] Riggall, A. C. & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. The Journal of neuroscience, 32(38), 12990-12998. DOI= https://doi.org/10.1523/JNEUROSCI.1892-12.2012

[17] Bettencourt, K. C. & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nature Neuroscience, 19(1), 150–157. DOI= http://doi.org/10.1038/nn.4174

[18] Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F. & Nyberg, L. (2015). Neurocognitive Architecture of Working

Memory. Neuron, 88(1), 33–46. DOI= http://doi.org/10.1016/j.neuron.2015.09.020

[19] Jonides J., Lewis R. L., Nee D. E., Lustig C. A., Berman M. G. in Moore K. S. (2008). The mind and brain of short-term memory. Annu Rev Psychol. 59, 193–224. DOI= http://doi.org/10.1146/annurev.psych.59.103006.093615

[20] Postle B. R. (2006). Working memory as an emergent property of the mind and brain. Neuroscience. 139, 23–38. DOI= http://doi.org/10.1016/j.neuroscience.2005.06.005

Podoba omejene racionalnosti in povratni učinki spreminjanja odločitvenih okolij

The Image of Bounded Rationality and Feedback Effects of Modifying Choice Environments

Toma Strle
Center za kognitivno znanost
Pedagoška fakulteta Univerze v Ljubljani
Kardeljeva ploščad 16
1000 Ljubljana, Slovenija
+386 1 5892 200
toma.strle@pef.uni-lj.si

POVZETEK

V članku predstavim idejo omejene racionalnosti, empirični program hevristik in pristranosti ter nekatere nadaljnje izsledke vedenjskih znanosti o odločanju, ki kažejo, da so odločitve ljudi neoptimalne in velikokrat pristrane. Nato na kratko predstavim področje odločitvenih spodbud, ki poskuša s spremembami odločitvenih okolij in podobnimi intervencijami človeško odločanje izboljšati. Ustavim se tudi pri vedno bolj prisotni manipulaciji odločanja s pomočjo "pametnih" algoritmov in velikih podatkov. V zadnjem delu prispevka se sprašujem o možnih povratnih učinkih spreminjanja odločitvenih okolij. Vedenjske znanosti o odločanju in programi poseganja v odločitvena okolja namreč ne upoštevajo možnosti, da posegi v odločitvena okolja ne spreminjajo le posameznih odločitev, ampak tudi odnos odločevalcev do odločanja in s tem morebiti spremenijo tudi odzive odločevalcev na same posege. Posegi v odločitvena okolja tako potencialno povratno vplivajo sami nase, tj. na to, kako na odločevalce učinkujejo.

Ključne besede

Hevristike, odločanje, odločitvene spodbude, omejena racionalnost, pristranosti, povratni učinki, spremembe odločitvenih okolij.

ABSTRACT

First, I present the idea of bounded rationality, the heuristics and biases research programme, and some further findings of behavioural decision sciences which show that people's decisions are not optimal and are many times biased. Next, I briefly present the decision nudge programme that aims at improving human decision-making by modifying choice environments. I also briefly stop at choice manipulation by "smart" algorithms and big data. In the last part of the paper, I consider possible circular effects of modifying choice environments. I argue that decision sciences and various attempts at modifying choice environments do not take into account the possibility that changes in choice environments not only affect a certain range of choice but also attitudes of decision-makers towards decision-making. However, by that, they also alter how decision-makers react to the implemented interventions and changes. Changes in choice environments thus potentially exert influence on themselves; i.e., they exert influence on their desired effects on decision-makers.

Keywords

Biases, bounded rationality, choice environment modification, decision-making, decision nudge, heuristics, feedback effects.

1. OMEJENA RACIONALNOST

Herbert Simon je v svojih seminalnih delih [20, 21] podal odmevno kritiko idealizirane podobe odločanja, ki so jo zagovarjale takratne normativne teorije. Slednje so v grobem predpostavljale, da so odločevalci zelo dobro informirani, da natančno poznajo svoje preference, izide potencialnih odločitev, da so izidom sposobni pripisovati vrednosti ali (subjektivne) koristi in da so, na podlagi takšnih podatkov, zmožni "izračunati" optimalno pot oz. pot, ki je vsaj blizu tega ideala.

Simon je takšno idealizirano podobo odločanja zamenjal z bolj realno, ekološko smiselno podobo. Zanj odločevalci ne posedujejo nekakšne vsevedne in računsko neomejene racionalnosti, ampak so v njej močno omejeni (od tod pojem omejena racionalnost; ang. *bounded rationality*). V članku *A Behavioral Model of Rational Choice* Simon npr. pravi: "Če posplošim, je naša naloga, da zamenjamo globalno racionalnost ekonomskega človeka z neke vrste racionalnim vedenjem, ki je kompatibilno z dostopom do informacij in računskimi zmožnostmi, ki jih organizmi, vključno z ljudmi, dejansko posedujejo v okoljih, v kakršnih ti organizmi obstajajo." [21, p. 99] V članku *Rational choice and the structure of the environment* nadalje razlaga: "Ker organizem […] nima ne čuta ne pameti, da bi odkril "optimalno" pot – če sploh domnevamo, da je koncept optimalnega jasno definiran –, se moramo ubadati le z iskanjem mehanizma izbire, ki ga bo vodil v zasledovanje "zadovoljive" poti; poti, ki bo dopuščala zadovoljitev vseh njegovih potreb na neki določeni ravni." [20, p. 136]

Simon skratka poudarja, da na odločanje ne moremo gledati kot na nekakšen proces optimizacije oz. proces racionalnega maksimiranja vrednosti ali koristi[1]. Prvič, zaradi omejenosti kognitivnega sistema; drugič, zaradi nedostopnosti vseh za odločitev relevantnih informacij; tretjič (to je moj dodatek), ker je

[1] Pomenljiva je Simonova opazka o nejasnosti definicije optimalnosti, ki jo lahko beremo kot dvom v načelno – vsaj normativno – opredeljivost optimalnosti izbir (glej tudi [23, 24]).

vprašljivo, če je v večini vsakodnevnih situacij čim boljša ali optimalna izbira sploh v interesu odločevalcev.

Simon je tako eden od začetnikov pogleda na odločanje, ki poskuša v zakup vzeti tako realnega odločevalca kot okolje, v katerem se ta odloča.

2. HEVRISTIKE IN PRISTRANOSTI TER NEKATERE DRUGE "ZMOTE" ODLOČANJA

Raziskovalni program hevristik in pristranosti[2] Tverskega in Kahnemana [31], njuna teorija odločanja pod pogoji tveganja (teorija obetov; ang. *prospect theory*) [13] in Simonove ideje o omejeni racionalnosti so močno zaznamovali sodoben pogled na presojanje in odločanje.[3] Le ta odločevalce koncipira kot organizme, ki jih vodijo predvsem nezavedne hevristike in ki se, vsaj v določenih kontekstih[4], v svojih presojah sistematično motijo ter sklepajo slabe, vsekakor pa ne optimalne ali zanje najboljše odločitve.

Tversky in Kahneman sta v svojem seminalnem članku iz leta 1974 [31] opisala tri intuitivne hevristike, nekakšne mentalne bližnjice, ki vodijo naše presoje. Glavna funkcija takšnih hevristik je poenostavljanje kompleksnosti problemov, ki v določenih okoliščinah vodi do pristranih presoj in odločitev: hevristiko reprezentativnosti (ang. *representativness heuristic*), hevristiko sidranja in prilagajanja (ang. *anchoring and adjustment heuristic*) ter hevristiko dostopnosti (ang. *availability heuristic*).

Naj kot primer opišem hevristiko sidranja in prilagajanja. Pri tej hevristiki so naše ocene raznih količin osnovane na začetnih vrednostih, ki jih, ko o neki stvari presojamo, nezadostno prilagodimo in tako podamo pristran odgovor. To se zgodi, čeprav so morda stvari, o katerih presojamo, z začetnimi vrednostmi povsem nepovezane. Različne podane začetne vrednosti tako vodijo v različne ocene enakega problema. V kontekstu odločanja so učinki te hevristike lepo razvidni v študiji Englicha, Mussweilerja in Stracka [5]. V svoji študiji so nemškim sodnikom s povprečno petnajst let izkušenj prebrali opis ženske, ki so jo ujeli pri kraji v trgovini. Nato so sodniki vrgli kocko, katere met je imel za rezultat vedno 3 ali 9. Ko se je kocka ustavila, so sodnike vprašali, ali bi žensko obsodili na čas zapora v mesecih, ki je večji ali manjši od številke na kocki. Potem so jih vprašali, na koliko mesecev bi jo obsodili. Ko je met kocke pokazal 9, je bila povprečna predlagana zaporna kazen osem mesecev, ko 3, pet mesecev. S tem so nakazali na možnost, da hevristike (v tem primeru hevristika sidranja in prilagajanja) vodijo tudi življenjsko pomembne odločitve.

Naj na kratko predstavim nekaj klasičnih pristranosti presojanja in odločanja za ponazoritev sodobne predstave o odločanju. V kontekstu odločanja je ena najbolj odmevnih "pristranosti" – ki predstavlja tudi primer kršitve enega izmed osnovnih aksiomov začetnih normativnih modelov odločanja – učinek uokvirjanja [13, 32]. Gre za to, da različne formulacije odločitvenega problema vplivajo na to, kaj izberemo (tudi zaradi dostopnosti različnih vidikov problema), ne glede na to, da imajo drugače formulirani izidi logično enako verjetnost dogoditve (klasičen primer je npr. problem azijske bolezni [32]). Nadalje, naše preference niso tako stabilne, kot so včasih mislili – študije kažejo, da so močno odvisne od naših predhodnih izbir [4, 6]. Študije tudi kažejo, da raje izbiramo manjše trenutne ali kratkoročne nagrade v primerjavi z večjimi, v času bolj oddaljenimi [17]. Nadalje, ljudje smo močno nagnjeni k temu, da ostajamo pri varnih izbirah oz. pri tem, kar že imamo (ang. *status quo bias*), tudi v primeru, ko aktivna izbira prinaša očitne dobičke [1]. Čustva, ki so sicer bistvena za odločanje, velikokrat vodijo do slabih izbir [1, 2]. Vsaj v nekaterih kontekstih se slabo odločamo tudi z vidika lastne sreče [8]. Ne nazadnje, včasih smo celo slepi za lastne izbire – nimamo dobrega vpogleda v razloge za lastne, dozdevno jasne in preudarne, odločitve [10].

Seveda našteti primer primeri specifičnih pristranosti in "zmot" odločanja kot taki ne pomenijo, da je odločanje človeka vedno ali večinoma zmotno oz. slabo. Vseeno pa se znanosti o odločanju bolj nagibajo k pogledu, da ljudje nimamo prav veliko vpliva na lastne odločitve, ki so obenem – pa naj si bodo sklenjene preudarno ali intuitivno – velikokrat podvržene najrazličnejšim pristranostim.

Na podlagi takšne podobe odločanja se nekateri raziskovalci zavzemajo za implementacijo specifičnih intervencij, ki bi posameznikom in/ali družbi kot celoti omogočile boljše odločanje in končne izbire. V nadaljevanju se bom osredotočil na strategijo spreminjanja odločitvenih okolij, predvsem na program odločitvenih spodbud.

3. SPREMINJANJE ODLOČITVENIH OKOLIJ

Eden izmed programov poseganja v odločitvena okolja, ki si za cilj postavlja izboljšati človeško odločanje, je program odločitvenih spodbud (ang. *decision nudge*[5]). Thaler in Sunstein [26, 29], ki sta program idejno zasnovala, zagovarjata t. i. dobronamerni libertarni paternalizem – dobronamerno vodenje človeških izbir z ohranjanjem svobode izbire: "Opremljen z razumevanjem vedenjskih izsledkov o omejeni racionalnosti in omejeni samokontroli, bi moral libertarni paternalist poskušati voditi človeške izbire v smeri spodbujanja blaginje, ne da bi odpravil svobodo izbire." [26, p. 1159]

Program odločitvenih spodbud kot eno izmed poti spreminjanja odločitvenih okolij predlaga spremembo privzetih pravil oz. izbir (med drugim na podlagi spoznanj o pristranosti statusa *quo* [1] in averzije do izgub [13]). Predlaga, da v odločitvena okolja vgradimo takšna privzeta pravila oz. privzete začetne/avtomatične izbire, ki so za odločevalce dobre oz. boljše kot te, ki so trenutno prisotne – predvsem takšne, ki povečujejo premoženje, blagostanje in zdravje posameznika oz. družbe. Ljudje se za aktivno izbiranje namreč ne odločajo prav pogosto, če pa že se oz.

---

[2] Ožje gledano pristranost pomeni sistematično napako, npr. v presojanju.

[3] Omenjena članka sta najbolj citirana članka s področja presojanja in odločanja (glede na pregled člankov, identificiranih po naslednjih ključnih besedah v naslovu na portalu Web of Science: decision* ali decid* ali choice* ali choos* ali judg* ali risk* ali uncertain* ali heuristic* ali bias*).

[4] Čeprav Kahneman [12] zagovarja stališče, da je preučevanje pristranosti skladno s pogledom na intuitivno mišljenje in odločanje kot v splošnem uspešno, pa bi po mojem večino raziskovalcev presojanja in odločanja, ki sledijo programu hevristik in pristranosti, trdilo, da je takšnih kontekstov pravzaprav ogromno.

[5] *Nudge* bi lahko prevajali tudi z dregljaj ali sunek; sam bom uporabljal besedo odločitvena spodbuda, saj gre, vsaj v osnovi, za dobronamerne posege.

so v to spodbujeni, trdi program, se mnogokrat ne odločajo sebi v prid. Dobre privzete izbire so z vidika programa odločitvenih spodbud tako večinokrat boljša alternativa od odločitvenih okolij, kjer se morajo odločevalci aktivno odločati. Na primer (primeri so vzeti iz [28]):

- Če želimo, da uporabniki menz jedo bolj zdravo, jim v prostoru restavracije na primer zdrave izdelke predstavimo pred nezdravimi; ne pa, da uporabnikom predstavimo obe vrsti hrane skupaj in jih pozovemo, naj se aktivno odločijo za zdravo hrano.

- Če želimo, da ljudje (več) varčujejo za pokojnine, spremenimo odločitveno okolje tako, da jih avtomatično vpišemo v določeno varčevalno shemo (še bolje je, da se v shemi mesečni znesek varčevanja povečuje z rastjo dohodka); ne pa obratno, da se morajo aktivno vpisati v varčevalno shemo, da sploh varčujejo.

- Če želimo, da ljudje po smrti darujejo svoje organe, jih ob rojstvu avtomatično opredelimo kot darovalce organov ali pa se morajo kot pogoj za pridobitev vozniškega dovoljenja opredeliti do tega ali želijo po smrti darovati organe (slednje je sicer delno že primer aktivnega odločanja).

Znanje o človeškem vedenju z namenom izboljševanja družbe in odločitev posameznikov uporablja tudi širše področje t. i. vedenjskih uvidov (ang. *behavioural insight*). Spekter uporabe je zelo raznolik: spoznanja se uporablja kot vodilo pri analizi, spreminjanju in ustvarjanju družbenih politik (*policy-making*) in struktur; kot vodilo oz. pomoč pri analizi in reševanju ekonomskih problemov; za namen spodbujanja specifičnih odločitev itd. Takšnih strategij za modifikacijo družbenih struktur in vplivanja na odločevalce se vedno bolj poslužujejo tudi mnoge inštitucije, vlade in korporacije: npr. Evropska komisija, Evropska unija, lokalne vlade, Svetovna banka [14, 16, 30, 35].

V literaturi je sicer zaslediti vedno več diskusij o tem, ali dobronamerno – oz. tako vsaj eksplicirano – spreminjanje odločitvenih okolij pravzaprav pomeni neupravičeno manipulacijo odločanja in vedenja [3, 11, 27]. Ne glede na to, kako se do vprašanja opredelimo, pa uvid v odločanje ljudi odpira mnogotere možnosti zlonamerni manipulaciji odločanja oz. manipulaciji odločanja v smereh, za katere si lahko predstavljamo, da si jih ljudje ne bi želeli. Primer je vedno bolj prisotna manipulacija odločanja s pomočjo "pametnih" algoritmov, velikih podatkov in uvidov v osebnost posameznikov: od manipulacije potrošniških izbir in izbir na volitvah do manipulacije vedenja uporabnikov spleta, socialnih omrežij in raznih aplikacij [15, 36, 37].

S temi zelo pomembnimi vprašanji se v nadaljevanju ne bom več ukvarjal. Posvetil se bom možnosti, ki jo – tako dobronamerne kot slabonamerne – intervencije spreminjanja odločitvenih okolij ne upoštevajo. Na kratko bom orisal možnost povratnih učinkov spreminjanja odločitvenih okolij na same učinke intervencij, ki naj bi odločanje "potisnile" v to ali drugo smer.

4. POVRATNI UČINKI SPREMINJANJA ODLOČITVENIH OKOLIJ

Znanosti o odločanju in programi poseganja v odločitvena okolja ne upoštevajo možnosti, da posegi v odločitvena okolja ne spreminjajo le posameznih odločitev, ampak tudi same odločevalce in njihov odnos do odločanja: na primer njihova implicitna ali eksplicitna prepričanja o odločanju, voliciji ali samokontroli; njihove motive za odločanje; koliko (aktivnega) odločanja si odločevalci sploh želijo; odločevalčeve predstave o lastnih sposobnostih, ki so relevantne za odločanje; njihovo vrednotenje smiselnosti tehtnega razmisleka o odločitvah. [glej tudi 22, 25 za podobno idejo v drugih kontekstih]. Posegi v odločitvena okolja tako inherentno odpirajo možnost, da se odločevalci na posege skozi čas začno odzivati drugače kot so predvidevali "arhitekti" odločitvenih okolij. Kajti odločevalci prav zaradi posegov – in podobe omejenosti odločanja, ki jo ti implicitno vnašajo v odločitvena okolja – potencialno spremenijo lasten odnos do odločanja in tako morebiti tudi to, kako se odločajo. Posegi v odločitvena okolja tako potencialno povratno vplivajo sami nase, tj. na to, kako na odločevalce skozi čas učinkujejo (ideja delno izvira iz del Hackinga [npr. 7] in Varele [npr. 33]).

Johnson in sodelavci [11, p. 488-490] se v svojem članku med drugim sprašujejo o tem, koliko alternativ naj "arhitekt izbire" predstavi potencialnemu odločevalcu. Arhitekt, pravijo avtorji, je v svoji odločitvi soočen z različnimi vprašanji. Je določeno število predstavljenih alternativ prenizko ali previsoko? Naj vse alternative predstavi naenkrat ali eno za drugo? V kakšnem vrstnem redu naj jih predstavi? Pravijo, da mora arhitekt izbire pri odgovoru upoštevati in najti ravnotežje med dvema kriterijema: a) več predstavljenih možnosti po eni strani pomeni večjo verjetnost, da odločevalcu ponudimo možnost, ki se ujema z njegovimi preferencami; b) po drugi strani več možnosti pomeni večjo kognitivno obremenitev. Da lahko arhitekt odgovori na to vprašanje ravnotežja, mora po njihovem v zakup vzeti tudi značilnosti odločevalcev: a) koliko se je odločevalec pripravljen ukvarjati s procesom izbire; b) zadovoljstvo odločevalca s procesom odločanja; c) bolj splošno, značilnosti procesov, ki vodijo do končne izbire; d) dodajajo še, da je odgovor na vprašanje ravnotežja odvisen tudi od lastnosti posameznih odločevalcev (npr. starost).

Vse te značilnosti odločevalcev so vsaj delno odvisne od specifičnih, kulturno pogojenih, odločitvenih okolij, s katerimi so odločevalci v interakciji, v katerih živijo in se v njih odločajo. Študije kažejo, da je pomen osebne izbire (da se za nekaj lahko odločimo sami, namesto da odločitev za nas sklene nekdo drug) in motivacija, ki iz nje izhaja, močno odvisna od kulturnega okolja [9]. Enako velja za zadovoljstvo z odločanjem [18] ali prepričanja o mentalnem naporu in iz njih sledečo sposobnost izvajanja samokontrole (ki predstavlja pomemben proces v odločanju) [19]. Na naše odločitve in dejanja vplivajo celo abstraktna prepričanja (npr. prepričanja o svobodni volji) [34], ki so lahko bistveno kulturno zaznamovana.

Podoben razmislek lahko naredimo o intervencijah, ki gradijo na prepričanju, da so dobre privzete izbire skoraj vedno boljše kot aktivno odločanje [28]. Res je, da smo ljudje kognitivno omejeni in da lahko preveč odločanja vodi v najrazličnejše negativne posledice za odločevalce. Po drugi strani ni jasno, kakšne posledice bi za odločanje (in učinke sami intervencij) prinesla družba, kjer bi dobronamerni vladar (arhitekt izbire) kreiral večino aspektov odločitvenih okolij. Morda bi "vseprisotno" zmanjševanje spodbude za aktivno, premišljeno in samoreflektirano odločanje vodilo v predrugačenje družbenih vrednot, kot sta recimo avtonomija odločanja in odgovornost za lastne odločitve. Morda bi imelo radikalno zmanjšanje spodbujanja aktivnega odločanja za posledico, da odločevalci ne bi imeli ne motivacije, ne veščine aktivnega odločanja, kar bi morda še povečalo možnost zlonamernih manipulacij odločanja s strani dozdevno dobronamernega arhitekta izbire. In čeprav je arhitektura izbire vedno prisotna – ne glede na to ali jo nekdo namerno ustvari ali ne –, pa spremembe odločitvenega okolja potencialno vodijo do spremembe samih odločevalcev in s tem do učinka, ki naj bi ga nanj imele.

5. ZAKLJUČEK

Trend sodobne družbe se pomika v smer relativno velikih sprememb odločitvenih okolij: vedno več inštitucij uporablja izsledke vedenjskih znanosti o odločanju za kreiranje in spreminjanje družbenih in bolj specifičnih odločitvenih okolij in vedno več podjetij poskuša odločanje manipulirati s pomočjo "pametnih" algoritmov, velikih podatkov in poznavanjem človeške duševnosti. Trend, ki odpira širok prostor, v katerem spremembe odločitvenih okolij – *na trenutno nepoznan način* – spremenijo same odločevalce, njihov odnos do odločanja in učinke, ki naj bi jih na odločevalce imele.

Posegi v odločitvena okolja v tem smislu na dolgi rok ne učinkujejo tako kot predvidevajo raziskovalci odločanja ali arhitekti izbire. Čeprav so odločevalca na začetku želeli spoznati "takšnega, kot je", z namenom, da bodo dosegli želene spremembe, so ga v svoji interakciji z njim že spremenili. Ob tem pa so potihoma pozabili, da odločevalec ni nespremenljiva entiteta, ki je neodvisna od interakcij z arhitektom izbire.

6. REFERENCE

[1] Anderson, C. J. 2003. The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin* 129, 1, 139-167. DOI= http://dx.doi.org/10.1037/0033-2909.129.1.139.

[2] Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. 2011. Do conscious thoughts cause behavior? *Annual Review of Psychology* 62, 1, 331-361. DOI= https://doi.org/10.1146/annurev.psych.093008.131126.

[3] Bovens, L. 2009. The Ethics of Nudge. In *Preference Change*, T. Grüne-Yanoff, & S. O. Hansson, Eds. Springer, Dordrecht, 207-219. DOI= https://doi.org/10.1007/978-90-481-2593-7_10.

[4] Brehm, J. W. 1956. Postdecision Changes in the Desirability of Alternatives. *Journal of Abnormal Psychology 52*, 3, 384-389. DOI= http://dx.doi.org/10.1037/h0041006.

[5] Englich, B., Mussweiler, T., & Strack, F. 2006. Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making. *Personality and Social Psychology Bulletin* 32, 2, 188-200. DOI= https://doi.org/10.1177/0146167205282152.

[6] Gerard, H. B., & White, G. L. 1983. Post-decisional Reevaluation of Choice Alternatives. *Personality and Social Psychology Bulletin* 9, 3, 365-369. DOI= https://doi.org/10.1177/0146167283093006.

[7] Hacking, I. 1995. The looping effect of human kinds. In *Causal cognition: A multidisciplinary debate*, D. Sperber, D. Premack, &. A. J. Premack, Eds. Clarendon Press, Oxford, 351-383, DOI= http://dx.doi.org/10.1093/acprof:oso/9780198524021.003.0012.

[8] Hsee, C. K., & Hastie, R. 2006. Decision and experience: why don't we choose what makes us happy? *Trends in Cognitive Sciences* 10, 1, 31-37. DOI= https://doi.org/10.1016/j.tics.2005.11.007.

[9] Iyengar, S., & Lepper, M. R. 1999. Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation. *Journal of Personality and Social Psychology* 76, 3, 349-366. DOI= https://doi.org/10.1037/0022-3514.76.3.349.

[10] Johansson, P., Hall, L., Sikström, S., & Olsson, A. 2005. Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310, 5745, 116-119. DOI= https://doi.org/10.1126/science.1111709.

[11] Johnson, E. J., Shu, S., Dellaert, B. G. C., Fox, C. R., Goldstein, D. G., Haeubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., & Weber, E. U. 2012. Beyond Nudges: Tools of a Choice Architecture. *Marketing Letters* 23, 487-504.

[12] Kahneman, D. 2003. A perspective on judgment and choice: mapping bounded rationality. *The American Psychologist* 58, 9, 697-720. DOI= https://doi.org/10.1037/0003-066X.58.9.697.

[13] Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2, 263-291.

[14] Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X. 2016. *Behavioural insights applied to policy: European Report 2016*. EUR 27726 EN. DOI= https://doi.org/10.2760/903938.

[15] Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. *PNAS* 114, 48, 12714-12719. DOI= https://doi.org/10.1073/pnas.1710966114.

[16] OECD. 2017. *Behavioural Insights and Public Policy: Lessons from Around the World*. OECD Publishing, Paris. DOI= https://doi.org/10.1787/9789264270480-en.

[17] Rachlin, H., Raineri, A., & Cross, D. 1991. Subjective probability and delay. *Journal of the Experimental Analysis of Behavior* 55, 2, 223-244. DOI= https://doi.org/10.1901/jeab.1991.55-233.

[18] Roets, A., Schwartz, B., & Guan, Y. 2012. The tyranny of choice: a cross-cultural investigation of maximizing-satisficing effects on well-being. *Judgment and Decision Making* 7, 6, 689-704.

[19] Savani, K., & Job, V. 2017. Reverse Ego-Depletion: Acts of Self-Control Can Improve Subsequent Performance in Indian Cultural Contexts. *Journal of Personality and Social Psychology* 113, 4, 589-607. DOI= https://doi.org/10.1037/pspi0000099.

[20] Simon, H. A. 1956. Rational Choice and the Structure of the Environment. *Psychological Review* 63, 2, 129-138. DOI= http://dx.doi.org/10.1037/h0042769.

[21] Simon, H. A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69, 1, 99-118.

[22] Strle, T. 2018. Looping minds: How cognitive science exerts influence on its findings. *Interdisciplinary Description of Complex Systems* 16, 4, 533-544. DOI= https://doi.org/10.7906/indecs.16.4.2.

[23] Strle, T. 2016a. Embodied, enacted and experienced decision-making. *Phainomena* 25, 98-99, 83-107.

[24] Strle, T. 2016b. Odločanje: od laboratorija do resničnosti vsakdanjega življenja. *Analiza* 20, 1, 61-84.

[25] Strle, T. & Markič, O. 2018. Looping effects of neurolaw and the precarious marriage between neuroscience and the law. *Balkan Journal of Philosophy* 10, 1, 17-26. DOI= https://doi.org/10.5840/bjp20181013.

[26] Sunstein, C., & Thaler, R. 2003. Libertarian paternalism. *American Economic Review* 93, 2, 175-179. DOI= https://doi.org/10.1257/000282803321947001.

[27] Sunstein, C. 2015. The Ethics of Nudging. *Yale Journal on Regulation* 32, 2, 413-450.

[28] Sunstein, C. R. 2017. Default Rules Are Better Than Active Choosing (Often). *Trends in Cognitive Sciences* 21, 8, 600-606. DOI= https://doi.org/10.1016/j.tics.2017.05.003.

[29] Thaler, R. H., & Sunstein, C. R. 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press, New Haven & London.

[30] Troussard, X., & van Bavel, R. 2018. How Can Behavioural Insights Be Used to Improve EU Policy? *Intereconomics* 53, 1, 8-12. DOI= https://doi.org/10.1007/s10272-018-0711-1.

[31] Tversky, A., & Kahneman, D. 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science* 185, 4157, 1124-1131. DOI= https://doi.org/10.1126/science.185.4157.1124.

[32] Tversky, A., & Kahneman, D. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211, 4481, 453-458. DOI= https://doi.org/10.1126/science.7455683.

[33] Varela, F. J. 1984. The creative circle: sketches on the natural history of circularity. In *The invented reality: Contributions to constructivism*, P. Watzlawick, Ed. Norton Publishing, New York, 309-325.

[34] Vohs, K. D., & Schooler, J. W. 2008. The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating. *Psychological Science* 19, 1, 49-54. DOI= https://doi.org/10.1111/j.1467-9280.2008.02045.x.

[35] World Bank. 2015. *World Development Report 2015: Mind, Society, and Behavior*. World Bank, Washington, DC. DOI= https://doi.org/10.1596/978-1-4648-0342-0.

[36] Youyou, W., Kosinski, M., & Stillwell, D. 2015. Computer-based personality judgments are more accurate than those made by humans. *PNAS* 112, 4, 1036-1040. DOI= http://dx.doi.org/10.1073/pnas.1418680112.

[37] Zuboff, S. 2015. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology* 30, 1, 75-89. DOI= https://doi.org/doi:10.1057/jit.2015.5.

Expected human longevity

Beno Šircelj
"Jožef Stefan" Institute
Jamova cesta 39
Ljubljana, Slovenia
beno.sircelj@gmail.com

Laura Guzelj Blatnik
"Jožef Stefan" Institute
Jamova cesta 39
Ljubljana, Slovenia
laura.g.blatnik@gmail.com

Ajda Zavrtanik Drglin
"Jožef Stefan" Institute
Jamova cesta 39
Ljubljana, Slovenia
ajda.zavrtanik@gmail.com

Matjaž Gams
"Jožef Stefan" Institute
Jamova cesta 39
Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT
In this paper we discuss the fall of ancient civilizations and possible future extinction causes for humans. Then we estimate the longevity of human civilization based on the absence of observable extraterrestrial civilizations and astronomical data using the Drake equation. If there are not many advanced civilizations in our galaxy, our longevity can be estimated at up to 10 000 years.

Keywords
Human extinction, Drake equation, Civilization collapse

1. INTRODUCTION
It seems that nothing in this world can last forever. For example, our Sun burns 600 million tons of hydrogen per second, generating the light that is required to make our planet habitable. According to astronomers, in 4-5 billion years it will go through an exciting, yet terrifying death, probably swallowing the Earth in the process. On the other hand, humans are familiar with the fact that nobody has yet lived over 122 years, but again find it hard to accept that nations and countries face similar destiny. When presented in the National Council of Slovenia that with the fertility rate below 1.6 Slovenians and other major nationalities in Slovenia will cease to exist in a couple of hundred years, the social media protests seemingly related to man-woman issues appeared, although the event was centered around the longevity issues [5]. But history teaches us that in the past there were many flourishing developed civilizations, yet none of them survived for a very long period of time. As people rise and fall, so do civilizations. The same is valid for human civilization – it will inevitably die out one day.

What are the possible causes of the end of our civilization? Using the Drake equation, can we predict how much time we have left and how can we extend it?

Finally, are we like a child cognitively incapable of accepting the incoming fate? Children understand the meaning of death only at the age of 5 [11]. Will we as a civilization live healthy and long or perish at first major obstacle?

1.1 Definition of a civilization
A civilization is defined as a complex society which is characterized by urban development, social stratification imposed by a cultural elite, symbolic systems of communication and a perceived separation from and domination over the natural environment [14].

1.2 Fall of ancient civilizations
There are many different reasons why ancient civilizations went into decline or even extinction. Here we mention a couple of civilizations and the reasons for their decline.

- **The Maya civilization**
  There are several theories about the fall of Maya civilization but the prevalent is that climate changes and consequent drought were the main causes. Other possible causes include social disorder, over-population and warfare [2].
- **Minoan Civilization**
  Minoans were one of the first civilizations in Europe. They were located in Crete, Greece and were wiped out by tsunamis following a volcanic eruption [4].
- **Roman Empire**
  Roman civilization probably collapsed due to many reasons. A weak army, constant barbarian invasions, political instability, overpopulation and epidemics are only few of the causes that might have lead to downfall of this once powerful civilization [12].
- **Native Americans**
  The decline of Native Americans happened when Europeans discovered America in 1492. They brought new diseases to the continent which no Native American was immune to. Furthermore, Europeans started with colonization and Native Americans had no proper weapons to resist them [15].

A recent theory from [8] claims that the downfall of most civilizations was accompanied by ideologies that conflicted with the production process due to changed conditions. The

stories of civilization declines are, however, often presented as cautionary tales to frighten us into correcting the error of our ways to prevent the end of our own global civilization [8]. They focus on climate change, human-caused environmental impacts and overpopulation because these three factors are the major global concerns of our time. They have a strong appeal to us because of the ubiquity of disaster-based stories. There are also several positive components in these stories, e.g. they promote environmental responsibility, global concern and sustainable growth.

Space or earth phenomena can cause extinction or at least significantly decrease the number of humans, as the Toba supereruption around 70 000 years ago indicates. It is estimated that at most 10 000 people survived at some point afterwards [9]. But these events might be less likely in the near future due to long intermediate intervals between Earth catastrophes and the relative short-term predictions of the human civilization [1], [10]. In this paper we concentrate on the human-induced problems.

While the doomsayers are a constant phenomenon in our life, and they come indeed in all forms and ideas, more often than not unsupported by data [6], serious analyses were often able to predict the human-caused grim outcomes. The scientific warnings should not be perceived as a pessimistic or doomsayer viewpoint, but as cautions to prevent major problems and even the collapse of our civilization. Whatever the case, with current knowledge and technology it might be possible to scientifically correctly predict most likely current civilization dangers and at least some estimations for the time-span of our civilization.

1.3 Possible causes of extinction of the human race

In this section we will discuss some of the possible causes for human extinction, based on an article by Bostrom [1].

- **Nuclear holocaust**
  USA and Russia hold about 93 percent of all nuclear weapons, but other countries are starting to stockpile them as well. Even worse, the treaty preventing arms races has been called off.

  There are various opinions whether an all-out nuclear war could eradicate humankind. Some believe that it would be hard to reach all possible settlements, for example the ones that are isolated from other people like the mountains of Tibet or remote islands in the South Pacific [13]. Also, there are nuclear shelters preventing the chosen ones. But even if we survive the initial impact, the long term climatic effects would lead to a nuclear winter and the number of people would decrease to drastically low numbers. While humans as a species would probably survive, the level of human civilization would decline dramatically, probably demanding several centuries to regain the former technological state – if ever.

- **Global warming**
  Ever increasing releases of greenhouse gasses could start a feedback loop and the temperatures could continue to rise. Even more species would go extinct and we would be unable to produce crops. If a negative spiral would enhance heating up the planet, life would become unbearable outside with negative consequences similar to those of a nuclear winter.

- **Artificial intelligence**
  The development of artificial intelligence will likely lead to superintelligence in the future. It is possible that in the case of a conflict between humans and superintelligence the entire human civilization could get annihilated [16].

- **Pandemic**
  A new deadly disease could infect the entire world population. There could be an genetically engineered biological agent with long latency and high mortality. Those viruses could be released by a lunatic or spawned unintentionally [7].

- **Asteroid or comet impact**
  This is a very small risk, but if an object 100 km wide would collide with Earth all advanced life could perish. There have been multiple extinctions on Earth and at least some were caused by impacts from space. The best known is the one eliminating dinosaurs around 65 million years ago when an object about 10 to 50 kilometers in diameter hit the Yucatan peninsula in Mexico. As a consequence, around 75 percent of all plant and animal species went extinct.

- **Accidental or deliberate misuse of nanotechnology**
  It might be possible to construct bacterium-scale nanobots that are self replicating and can feed on organic matter. Such robots could ultimately eat or destroy the entire biosphere. This is one of the examples where humans construct a new mechanism capable of destroying civilization.

Can we avoid extinction of a particular nation or the human civilization in the first place? Clearly, the answer is no, and the real question is how long will human civilization persist, analogous to a question about a particular individual. In the next section we present a model that predicts the longevity of human civilization based on the Drake equation.

2. THE DRAKE EQUATION AND ESTIMATING THE LONGEVITY OF HUMAN CIVILIZATION

In 1961 Frank Drake proposed an equation for calculating the number of detectable civilizations in our galaxy at any given moment. The equation consists of several parameters [3]:

$$N = R_* f_p n_e f_l f_i f_c L,$$

where R_* is the rate of star formation per year, f_p is the fraction of stars with planets, n_e is the number of Earth-like (or otherwise habitable) planets per star that has planets, f_l is the fraction of habitable planets with actual life, f_i is the fraction of life-bearing planets that develop intelligence, f_c is the fraction of intelligent civilizations that are detectable and L is the average longevity of such civilizations. Finally, N is the number of detectable civilizations. In the original

article the authors used point values to estimate each one of the parameters. Sandberg et al. used a different approach in [10] - instead of using point values they used probability distributions for the parameters listed on table 1.

We used the Drake equation with the Sandberg's approach for the basis for our calculations. Since the parameters in the equation are all estimates we can solve equation for L and take N as a variable. The estimation of N can be somewhat limited from observations of our stellar neighbourhood. The equation for computing L is therefore as follows:

$$L = \frac{N}{R_* f_p n_e f_l f_i f_c} \tag{1}$$

2.1 Estimation of parameters

We used probability distributions to model each variable as in the paper presented by Sandberg [10]. For the distribution of the number of civilizations in our galaxy we set the lower bound at $N = 1$ and the upper bound at 10^4. The reason for this estimate is as follows: We have been trying in vain to get a signal from foreign civilizations even though quite extensive and expensive searches of the universe were performed. The search for extraterrestrial intelligence (SETI) is a collective term for scientific searches for intelligent extraterrestrial life. Various methods and approaches are used to detect signs of transmissions from civilizations on other planets, but most commonly monitoring of electromagnetic radiation is performed. The first scientific investigations began in the early 1900s, and focused international efforts have been going on since the 1980s. While some consider UFOs as a proof of foreign civilizations visiting us, there are no scientifically confirmed results so far. Since some projects were carried out using huge resources and time, that is a rather disturbing indication. Furthermore, it should be noted that this paper relies purely on the known and generally accepted scientific knowledge and UFOs are not part of it. If we therefore assume that detectable civilizations are evenly distributed throughout the galaxy, then there are at most 10 000, since otherwise we would have already observed one (radius of the galaxy is 10^5 light years while we can detect signals as far as 10^3 light years). Consequently, the range of L is theoretically from 10^{-2} to 10^{13}, i.e. from 3 days to ten trillion years.

Parameter	Distribution
R_*	log-uniform from 1 to 100
f_p	log-uniform from 0.1 to 1
n_e	log-uniform from 0.1 to 1
f_l	log-normal rate, described in paper[10]
f_i	log-uniform from 0.001 to 1
f_c	log-uniform from 0.01 to 1
N	point values: 1 to 10 000

Table 1: Probability densities for the parameters in Equation (1)

To estimate the longevity of human civilization, we did not model the parameter N with distributions. Instead we used multiple point-values as inputs to the equation. For example, suppose there are 1, 10, 100, 1 000, 10 000 civilizations in our galaxy now – what can we conclude about our longevity in that particular case?

Several hundreds of models either of different nature or of significantly different parameters were designed and tested, but here we present only one.

3. EXPERIMENTS

The computing was performed in a stochastic way: for a chosen N, a value of each parameter was randomly generated using the predefined probability density, and L was computed according to the Drake equation. The obtained probability distribution denotes the longevity of human civilization under chosen probability distribution for the given parameters and for the chosen N – the number of technologically advanced civilizations in our galaxy, i.e. the ones that transmit electromagnetic signals to space. From the obtained probability density, several derived graphs can be generated, e.g. the one in Figure 1.

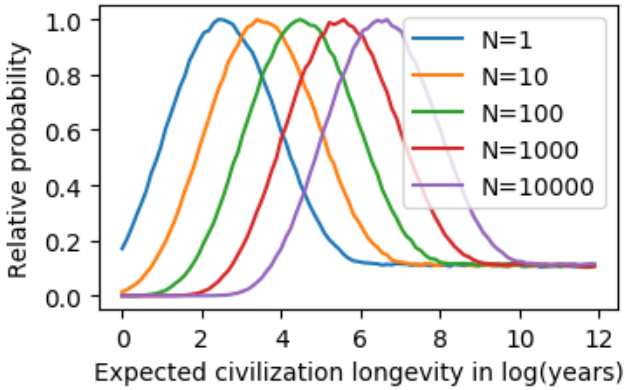


Figure 1: Graph for log(L), i.e. for expected human longevity based on the values of N – the number of civilizations in our galaxy.

N	median	stabilization	volume
1	2 200	13 600	2700
10	22 000	11 100	10 000
100	220 000	9 300	63 700
1000	2 200 000	5 800	545 600
10 000	22 000 000	/	1 000 800

Table 2: Median and stabilization values for different N.

The same relations are also presented in side-view in Figure 2 and in 3D in Figure 3. Bigger N seemingly corresponds to better chances for longer human longevity, in a positive correlation with N. In addition, our longevity is obviously limited, but the exact relations are somehow difficult to comprehend due to the non-linear scale.

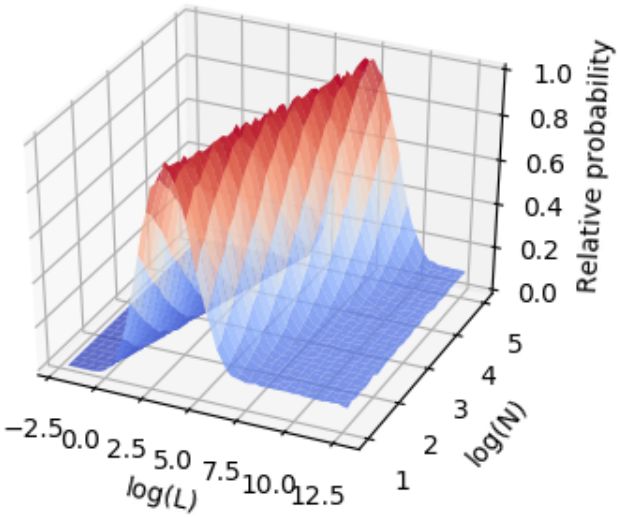


Figure 2: Longevity based on N, side view.

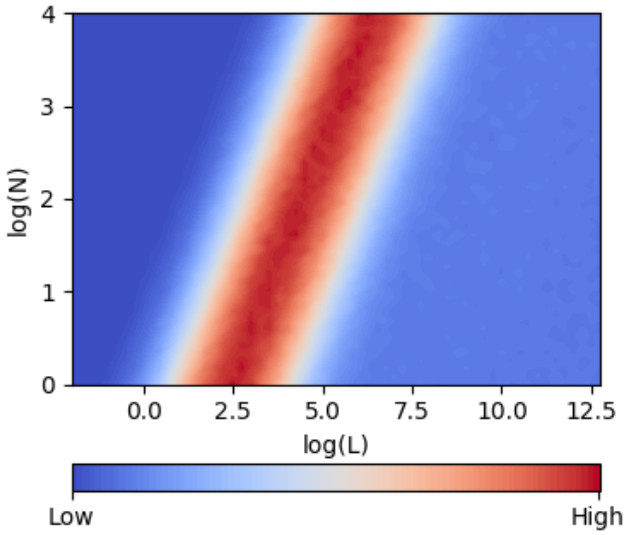


Figure 3: Longevity based on N, top view.

If instead of logarithmic scale, the graph of probability densities is presented in a linear scale (Figure 4), the impression is now quite different. The "true" relation between N and L is as follows:the majority of possibilities for smaller N are at the left part of the graph resulting in a bigger bump accompanied with a slower decline. The point of stabilization, i.e. when a decline is less than 1 percent in a corresponding 100 years is presented in Table 2 as "stabilization". One can also calculate median longevity by computing it for each graph, denoted as "median". The difference between "stabilization"

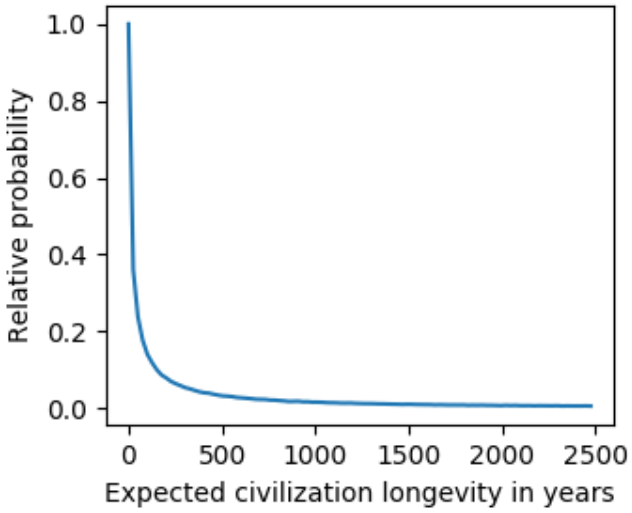


Figure 4: Graph for longevity, i.e. $\log(L)$ in linear scale for $N = 1$.

and "median" is that median represents a point dividing all simulations into two equally frequent intervals, while stabilization indicates the end of steep, i.e. more than 1 percent decline in the probability densities. While median linearly grows with the number of civilizations, stabilization declines denotes where the peak in probability densities on the left is getting smaller than 1 percent. At N equal to 10 000, no decline is bigger than 1 percent. The right-most column "volume" denotes the percentage of the current integral of probability densities in a millennium decreases to less than 1 percent compared to the best 100 years (normalized). These relations are highlighted in Table 3 where the average over an interval 0..1000, 1000..2000, 10 000..11 000 etc. is divided by an average over best (i.e. usually the first) 100 years. These numbers denote how much more probable are the first 100 years compared to the first 1000 etc.

N	0-1000	1000-2000	10 000-11 000	100 000-101 000	1 000 000-1 001 000
1	0.186	0.024	0.002	0.000	0.000
10	0.289	0.073	0.010	0.001	0.000
100	0.600	0.276	0.058	0.006	0.000
1000	0.871	0.789	0.298	0.053	0.005
10 000	0.275	0.749	0.843	0.299	0.048

Table 3: Probability densities of 1000 years for different N at 5 specific longevities normalized to the highest value in 100 years.

The reason why we present graphs in logarithmic scale is that the linear scale does not enable the reader to comprehend anything outside the relevant scope. For example, Figure 4 would consist of two lines if the max years would be 25 000 instead of 2 500 – one vertical on the left and one horizontal on the x axis.

4. CONCLUSION

The aim of this research was to establish probability densities of longevity of human civilization. In this paper we

presented results of just one model while we have tested hundreds of them. The model analyzed here shows that if there are more civilizations, we have higher probability of living longer. Regardless of N and after initial fluctuations very close to the left, the curve of longevity is monotonic, decreasing. At N equal to 1, i.e. if we are the only ones in our galaxy, we will probably live only for approximately 2 000 - 14 000 years. At N equal to 10, the expected high-probable longevity is from 11 600 to 22 000 years. At N equal to 10 000 there is no peak at the left and the probability density very slowly declines. In other words – there is not any explicit pattern and predictions are undecidable.

Our maximum survival time seems to be about 10 000 - 20 000 and maybe up to 100 000 years. But most likely, the expected time is substantially shorter.

This study might be relevant because it indicates that we need to start acting wisely sooner rather than later to prevent grim scenarios. Namely, if the predicted time would be say millions of years, there would be no need to go to Mars and other planets soon, we should not worry too much about global warming or other problems. But if the predictions indicate that these dangers might hamper our progress relatively quickly, at least in terms of cosmic timing, we should actively analyze them and react appropriately.

5. REFERENCES

[1] N. Bostrom. Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, 9(1), 2002.

[2] M. Cartwright. The Classic Maya Collapse. https://www.ancient.eu/article/759/, October 2014. [Online; accessed 30-August-2019].

[3] F. Drake. *The Drake Equation: Estimating the Prevalence of Extraterrestrial Life Through the Ages*. Cambridge University Press, 2015.

[4] W. L. Friedrich. The Minoan Eruption of Santorini around 1613 B.C. and its consequences. *Tagungen des Landesmuseums für Vorgeschichte Halle*, 2013.

[5] M. Gams. *Bela knjiga slovenske demografije*. Institut "Jozef Stefan" Ljubljana, 2019.

[6] S. Juster. *Doomsayers: Anglo-American Prophecy in the Age of Revolution*. University of Pennsylvania Press, 2006.

[7] G. McNicoll. Global trends 2015: A dialogue about the future with nongovernment experts. *Population and Development Review*, 27(2):385–385, 2001.

[8] G. D. Middleton. *Understanding collapse: Ancient history and modern myths*. Cambridge University Press, 2017.

[9] M. Rampino and S. Ambrose. *Volcanic winter in the Garden of Eden: The Toba supereruption and the late Pleistocene human population crash*, volume 345, pages 71–82. 01 2000.

[10] A. Sandberg, E. Drexler, and T. Ord. Dissolving the Fermi paradox. *arXiv preprint arXiv:1806.02404*, 2018.

[11] V. Slaughter. Young children's understanding of death. *Australian Psychologist*, 40, 11 2005.

[12] D. L. Wasson. Fall of the Western Roman Empire. https://www.ancient.eu/article/835/fall-of-the-western-roman-empire/, April 2018. [Online; accessed 30-August-2019].

[13] W. Wells. Apocalypse how? In *Apocalypse When?*, pages 93–128. Springer, 2009.

[14] Wikipedia contributors. Civilization Wikipedia, the free encyclopedia, 2019. [Online; accessed 17-September-2019].

[15] M. Wisniewski. The decline of native culture in america: Causes and effects. https://ahr-ashford.com/the-decline-of-native-culture-in-america-causes-and-effects-by-mark-wisniewski/, 2005. [Online; accessed 30-August-2019].

[16] E. Yudkowsky. Creating friendly ai 1.0: The analysis and design of benevolent goal architectures. *The Singularity Institute, San Francisco, USA*, 2001.

Indeks avtorjev / Author index

IS
2019

Konferenca / Conference

Uredili / Edited by

Kognitivna znanost / Cognitive Science

Toma Strle, Tine Kolenik, Olga Markič