Cognition and Artificial Life

2024

Proceedings

Cognition and Artificial Life 2024

Editors
Čeněk Šašinka
Zdeněk Stachoň
Stanislav Popelka
Michal Černý
Nicol Dostálová
David Košatka
Alžběta Šašinková

Fort Science,
Interactive Science Centre of Palacký University Olomouc, 15.5.–17.5. 2024

Cognition and Artificial Life 2024. Peer-reviewed proceedings.

MUNI ARTS — Department of Information and Library Studies

DEPARTMENT OF GEOINFORMATICS
Palacký University in Olomouc

Table of Contents

Navigating Classroom Challenges:
How VR and AI Could Elevates Teaching Strategies

Markéta Košatková[1], Jan Nehyba[2]

[1]Dept. of Social Education, Faculty of Education, Masaryk University, Brno, Poříčí 7/9, 603 00, Czech Republic
[2]Dept. of Education, Faculty of Education, Masaryk University, Brno, Poříčí 7/9, 603 00, Czech Republic
kosatkova@ped.muni.cz, nehyba@ped.muni.cz

Abstract

The presentation will introduce the pilot phase of the newly emerging project at the Faculty of Education of Masaryk University. The aim is to support student teachers and novice teachers in their efforts to gain competencies to manage challenging student behaviour in the classroom with future possible implications for parents and social change agents. Support will be provided through intervention tools that appropriately combine pedagogical principles and digital technologies.

The project leverages Artificial Intelligence (AI) algorithms and Virtual Reality (VR) to explore and implement successful case studies. The platform's AI-powered search enables teachers to find tailored recommendations on the problem behaviours that affect their students across various motivational to authoritative spectrums. At the same time, VR simulations provide immersive scenarios for practising these strategies.

This innovative approach enhances understanding of complex classroom dynamics and allows for practising communication strategies in a controlled, risk-free environment.

1 Bridging gaps: Innovations in Managing Classroom Behaviors

The field of classroom management is increasingly confronted with the challenge of disruptive behaviour in schools, a problem that research continues to underscore as both prevalent and complex to address (Wagner, 2020). The term "challenging behaviour" refers to conduct by children that markedly diverges from what is typically expected for their age, adversely impacting learning and interpersonal relationships within the school setting. Regardless of their experience level, teachers often find their skills and competencies inadequate for managing such behaviours, leading to professional dissatisfaction.

Moreover, there is a notable deficiency in systematic training for handling disruptive behaviours—a training that should ideally be grounded in empirical research and encompass a range of strategies for respectful communication (*e.g.,* Gordon, 2015; Nováčková & Nevolová, 2020; Rosenberg, 2012). A multidisciplinary approach that combines insights from education, psychology, computational linguistics, and interactive simulation presents a promising avenue for addressing this gap.

2 Tech-Enabled Teaching: AI and VR as Catalysts for Change

The project's goal is to nurture a culture of respectful communication and build positive interpersonal relationships, achieved through the training of educators at Masaryk University. This initiative is pivotal in enhancing well-being and upholding democratic values, bridging the divide between modern technological approaches and traditional educational practices to offer holistic support for today's educational challenges. The tools created as part of this project serve complementary functions and operate across three distinct intervention levels.

Firstly, the project addresses disruptive behaviour through a comprehensive database of case studies on challenging behaviours (*e.g.,* Fikarová & Škubalová, 2020). Given the noted lack of varied and relevant case studies in existing literature (Kohut & Range, 1986; Arum & Ford, 2012; Tal, 2010; Pullen & Mostert, 2010), there is a need for an extensive online repository. This platform aims to cover various cases of problem behaviour, offering educators practical and effective strategies.

Secondly, natural language processing (NLP) technology (Reimers & Gurevych, 2018) enhances our grasp of these case studies. It enables users to find or input case studies for detailed analysis using semantic search tools that examine the context and relationships within the data to provide a deeper understanding of behaviour patterns. This process facilitates the creation of personalized, AI-generated responses and solutions based on real case studies.

The final, most intensive tool for developing competencies in managing challenging classroom situations involves a virtual reality (VR) simulation environment (Mouw et al., 2020). VR dramatically increases the level of immersion, making the simulated classroom experiences feel real but in a risk-free environment (Lugrin et al., 2016). This immersive approach significantly enhances the training process, equipping educators to navigate the complexities of classroom management through somatic-cognitive experience more effectively.

We believe that these three tools—a vast database of case studies enhanced by AI and the immersive experience offered by VR—could significantly enrich the curriculum for teacher professional development, preparing them to manage challenging behaviours in the classroom more effectively.

3   The next step in supporting prospective teachers

We believe that these three tools—a vast database of case studies enhanced by AI and the immersive experience offered by VR—could significantly enrich the curriculum for teacher professional development, preparing them to manage challenging behaviours in the classroom more effectively.

Our objective is to enhance the skills of both aspiring and current educators, enabling them to address challenging behaviours constructively and respectfully right as they occur. This goal will be achieved by deploying supportive tools designed to augment, rather than replace, traditional classroom experiences and the systematic approaches developed by experts for various behavioural disorders. The key benefits we anticipate include:

- Personalized support: Utilizing an AI-driven platform to customize content (case studies and recommendations) according to the unique needs of users

- Enhanced Experiential Training: Leveraging immersive VR simulations to provide educators with realistic case studies for hands-on practice.

- Educator Training on Technological Integration: Offering methodology to teachers on effectively integrating AI and VR technologies into their teaching strategies.

The promise held by these technologies for classroom management and teacher training is substantial. However, realizing their full potential depends on their thoughtful application, continuous assessment, and the ability to adapt to the evolving requirements of both educators and students.

Literature

Arum, R., & Ford, K. (2012). How Other Countries" Do Discipline". *Educational Leadership, 70*(2), 56-60.

Fikarová, T., Škubalová, T. (2020) *Výukové kazuistiky k psychologickým předmětům společného základu* (Instructional Case Studies for Psychological Common Base Courses). MUNI. https://is.muni.cz/publication/1625818/en/Vyukove-kazuistiky-k-psychologickym-predmetum-spolecneho-zakladu/Fikarova-Skubalova

Gordon, T. (2015). *Škola bez poražených: praktická příručka efektivní komunikace mezi učitelem a žákem.* Malvern.

Kohut Jr, S., & Range, D. G. (1986). Classroom Discipline: Case Studies and Viewpoints. Publications Department, National Educational Association, Professional Library.

Lugrin, J. L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT, 3,* 26. https://doi.org/10.3389/fict.2016.00026

Mouw, J. M., Fokkens-Bruinsma, M., & Verheij, G. J. (2020). Using Virtual Reality to promote pre-service teachers' classroom management skills and teacher resilience: A qualitative evaluation. In *Proceedings of the 6th International Conference on Higher Education Advances (HEAd'20)* (pp. 325-332). Universitat. https://doi.org/10.4995/HEAd20.2020.11049

Nováčková, J., & Nevolová, D. (2020). *Respektovat a být respektován: cesta k sebeúctě a zodpovědnosti* (Přepracované vydání). PeopleComm.

Pullen, P. L., & Mostert, M. P. (2010*). Managing classroom behavior: A reflective case-based approach.* Prentice Hall.

Reimers, N., & Gurevych, I. (2018). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv*:1908.10084.

Rosenberg, M. B. (2012). *Nenásilná komunikace*. Portál.

Tal, C. (2010). Case studies to deepen understanding and enhance classroom management skills in preschool teacher training. *Early Childhood Education Journal, 38*(2), 143-152.

Wagner, J. (2020). Prohlášení školských odborů k uzavření škol. *Pedagogicke.info*. http://www.pedagogicke.info/2020/09/prohlaseni-skolskych-odboru-k-uzavreni.html?m=1

Seeing beyond borders: Exploring the world's diversity in visual perception

Jiří Čeněk

Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00, Brno, Czech Republic
144946@mail.muni.cz

Extended abstract

The aim of this speech is to give the audience an idea about development and state-of-the-art of the cultural and cross-cultural research on visual perception and cognition and should ultimately answer the question on how and how systematically members of various cultures differ in the way they perceive the visual world.

The contribution starts with a short overview of the history of the cultural research on visual perception citing the pioneering works by Hudson (1960), Deregowski (1980), or Segall et al. (1990). In these early works the authors generally focused on the comparisons of "civilized" or "modern" societies (as sometimes non-correctly labelled in early works; read USA, UK, Germany) with "primitive" or "traditional" societies (tribes and communities in Africa, Asia, or Pacific). Although theoretically important and interesting, these studies contained certain methodological flaws that by today standards allow only limited interpretability.

Research on cognitive styles, more specifically on holistic and analytic cognitive style (Nisbett & Masuda, 2003) tried to overcome these flaws by investigating cultural groups that are better comparable (e.g. more similar nature of surrounding physical environment, comparable level of economic development).

Dr. Čeněk and his research group from Masaryk University performed one of the largest (regarding the number of countries and the number of participants) cross-cultural experimental studies in the field of perception. A test battery composed of methods with various levels of complexity was created; from relatively simple Navon hierarchical figures (Navon, 2003) based tests, to relatively complex tasks that utilize eye-tracking and immersive virtual reality methodologies.

Based on this research several theoretical pitfalls present in the field of cross-cultural perception have been identified and several suggestions for further development of the discipline were formulated.

Literature

Deregowski, J. B. (1980). *Illusions, patterns and pictures*. Academic Press.

Hudson, W. (1960). Pictorial depth perception in sub-cultural groups in Africa. *The Journal of Social Psychology, 52*(2), 183-208.

Navon, D. (2003). What does a compound letter tell the psychologist's mind? *Acta Psychologica, 114*(3), 273-309.

Nisbett, R., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences, 100*(19), 11163-11170.

Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1990). *Human behavior in global perspective*. Allyn and Bacon.

Classification of Dyslexia using AI-based Methods Applied to Different Eye-Tracking Data Representations

Jan Sedmidubský[1], Dostálová Nicol[2]

[1]Department of Machine Learning and Data Processing, Faculty of Informatics, Masaryk University, Botanická 68A, 602 00 Brno, Czech Republic
[2]Department of Educational Sciences, Faculty of Arts, Masaryk University, Arne Nováka 1, 602 00 Brno, Czech Republic
sedmidubsky@mail.muni.cz, nicol.dostalova@mail.muni.cz

Extended abstract

Dyslexia is a specific learning disability characterized by severe difficulties with reading such as errors in decoding and spelling (Lyon et al., 2003). These symptoms subsequently lead to a decreased overall reading process and text comprehension (Peterson & Pennington, 2012) followed by a negative impact on the social and emotional development of an individual (Livingston et al., 2018). For this reason, it is crucial to focus on a detailed and accurate diagnosis of dyslexia, which would then enable the selection of an appropriate intervention approach.

One of the main research objectives – in the field of dyslexia diagnostics – focuses on the evaluation how much a given individual is prone to dyslexia. Traditional research (e.g. Rayner, 1998; Hutzler & Wimmer, 2004) shows that eye movements of dyslexic individuals differ from intact ones during the process of text reading. The differences are, for example, in the number of eye fixations, the fixation duration, or the saccadic amplitude. Since it is principally possible to record eye-tracking data and extract such eye movement features, the current research focuses on the application of artificial intelligence (AI) approaches to classify the dyslexic patterns automatically (Nilsson Benfatto et al., 2016; Smyrnakis et al., 2017).

The main goal of this keynote talk is to share our three-year experience with: (1) collecting a new eye-tracking dataset of dozens of subjects on selected text-reading tasks, (2) extracting various types of eye-tracking data representations, and (3) applying several AI-based classification methods to such representations. Specifically, we briefly introduce two kinds of data representations: visual-based representation that encodes information about eye fixations in the form of 2D image, and feature-based representation that encodes task-specific eye-movement characteristics in the form of a high-dimensional vector. We apply residual neural networks (He et al., 2016) to classify visual-based representations and *k*-nearest neighbor and Multi-Layer Perceptron (MLP) approaches to classify feature-based representations. The best-performing combinations can achieve the classification accuracy up to 90%.

Literature

He, K., Zhang, X., Ren, & S. J. Sun. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Las Vegas, USA, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language*, *89*, 235–242. https://doi.org/10.1016/S0093-934X(03)00401-2

Livingston, E. M., Siegel, L. S., & Ribary, U. (2018). Developmental dyslexia: Emotional impact and consequences, *Australian Journal of Learning Difficulties*, *23*(2), 107–135. https://doi.org/10.1080/19404158.2018.1479975

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A Definition of Dyslexia. *Annals of Dyslexia*, *53*, 1–14. https://doi.org/10.1007/s11881-003-0001-9

Nilsson Benfatto, M., Öqvist Seimyr, G., Ygge, J., Pansell, T., Rydberg, A., & Jacobson, C. (2016). Screening for Dyslexia Using Eye Tracking during Reading. *PLOS ONE*, *11*(12), e0165508. https://doi.org/10.1371/journal.pone.0165508

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *The Lancet, 379*(9830), 1997–2007. https://doi.org/10.1016/S0140-6736(12)60198-6

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Smyrnakis, I., Andreadakis, V., Selimis, V., Kalaitzakis, M., Bachourou, T., Kaloutsaki, G., Kymionis, G. D., Smirnakis, S., & Aslanides, I. (2017). RADAR: A novel fast-screening method for reading difficulties with special focus on dyslexia. *PLoS ONE, 12*(8), e0182597. https://doi.org/10.1371/journal.pone.0182597

Evaluating the Attractiveness of Fort Science Geographic Display Using Mobile Eye-tracking Glasses

Stanislav Popelka, Jiří Vysloužil

Department of Geoinformatics, Palacký University Olomouc
17. listopadu 50, 77146, Olomouc, Czech Republic
stanislav.popelka@upol.cz

Abstract

This study assesses the attractiveness of geographic displays through mobile eye-tracking technology. Early qualitative eye movement research began in the 18th century, with significant advancements by Javal (1879) and Lamare (1892) in observing eye movements during reading. A major breakthrough came when Hartridge and Thompson (1948) created the first mobile device that attached only to the participant's head. Modern eye-trackers are now available as small, lightweight glasses used in various fields such as sports, marketing, transportation, healthcare or cartography, with a recent increase in museum applications. The study aimed to evaluate the "Living Water" exhibit at the Fort Science, a science museum established in 2015 under the Faculty of Science, Palacký University in Olomouc. Fourteen interactive geography exhibits were selected for evaluation, ranging from puzzle-like exhibits to those with changing animations, like the flood model of Olomouc. The study involved museum visitors to maintain ecological validity, with 36 participants divided into children, students, and adults. Data were collected using Tobii Pro Glasses 3. Calibration quality was verified with GlassesValidator from Utrecht University. Data processing in Tobii Pro Lab involved annotating video recordings to identify when visitors engaged with exhibits, using "Snapshot" function for automatic exhibit recognition, creating areas of interest, generating attention maps, and analyzing eye-tracking metrics. Results categorized exhibits into attractive and less attractive, suggesting a new layout to increase observation time for less attractive exhibits. The study highlights the detailed methodology of collecting and analyzing data with mobile eye-tracking devices for evaluating interactive exhibits, offering potential applications for Fort Science.

Acknowledgements

References

Javal, E. (1878). Essai sur la physiologie de la lecture. *Annales d'Ocilistique*, *80*, 97-117.

Lamare, M. (1892). Des mouvements des yeux dans la lecture. *Bulletins et Mémoires de la Société Française d'Ophthalmologie*, *10*, 354-364.

Hartridge, H., & Thomson, L. C. (1948). Methods of investigating eye movements. *The British journal of ophthalmology*, *32*(9), 581.

Exploring the Relation between Reading Style and Cognitive Style: an Eye-Tracking Approach

Nicol Dostálová[1], Čeněk Šašinka[2], David Lacko[2]

[1]Department of Educational Sciences, Faculty of Arts, Masaryk University
[2]Department of Information and Library Studies, Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00, Brno, Czech Republic
nicol.dostalova@mail.muni.cz, cenek.sasinka@mail.muni.cz, david.lacko@mail.muni.cz

Extended abstract

The discussion surrounding the consistency of reading styles across various texts and reading tasks persists, leaving unanswered questions about their role in shaping individual reading styles and their connection to cognitive styles. In the context of different manners of reading, Rayner defined so-called risky and conservative reading styles based on the patterns of eye movements while reading various texts, based e.g. on number of regressions or word skipping (Rayner et al., 2006; Rayner et al., 2009).

However, it is still not yet fully evident how reading patterns and strategies are profiled and molded into reading styles, as well as how they relate to cognitive styles, which can be thought of as the way we think and solve problems, learn, etc. (Witkin et al., 1977).

The main aim of this paper is to present a research study (see Lu et al., 2023) in which we decided to implement an eye-tracking method to validate Rayner's concept of reading styles on a specific reading task. For this purpose, rudimentary reading stimuli were created during which adult participants read several selected texts divided into three paragraphs. Cognitive styles were subsequently assessed using the Compound Figure Test (CFT) (a PC version of the original Navon test; see Navon et al., 1977).

The results show that stability in eye movements was preserved during reading of the several text stimuli. No relationship between cognitive style and reading style was found for selected eye movement patterns during reading and cognitive style scores. In conclusion, these results may provide new insights into the way of handling and reading different textual materials.

Acknowledgements

Literature

Lu, W., Dostálová, N., Lacko, D., Šašinková, A., & Šašinka, Č. (2023). From reading style to cognitive style and its possible application: An Eye-Tracking and CFT approach. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1). https://doi.org/10.1186/s40862-023-00195-4

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging*, *24*(3), 755–760. https://doi.org/10.1037/a0014300

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, *47*(1), 1–64. https://doi.org/10.2307/1169967

The effect of word frequency and orthographic neighbourhood density on reading in people with dyslexia

Monika Ptáčková

Faculty of Education, Palacký University Olomouc
Žižkovo náměstí 5, Olomouc, 779 00, Czech Republic
monika.ptackova01@upol.cz

1 Introduction

Reading is one of the skills used in everyday life in the current population. People with developmental reading disorders (dyslexia) struggle with reading despite of good motivation and adequate teaching of this skill. Many studies have shown that eye movements differ in people with developmental reading disorder (dyslexia) compared to typical readers (see De Luca et al., 2002; Jošt, 2009; Vyhnálek et al., 2006 and others).

In a theoretical frame of the dual route model of reading (Coltheart, 2006) differences in eye movements during word recognition depended on specific word properties could point out the differences in using of the lexical and sublexical ways of reading. Specifically word frequency effect[1] is the indicator of lexical way whereas the orthographic neighbourhood density effect[2] can according to some authors be an indicator of sublexical reading (Laxon et al., 2002). However, the effect of the orthographic neighbourhood in reading is still at the centre of discussion (Guitard et al., 2024; Nemati et al., 2022 and others).

Furthermore, these effects and their differences can be related to different characteristics of the language itself and the reading process as well as its specifics can differ not only because of the presence/absence of reading disorder but also because of the language characteristics as such (Suárez-Coalla & Cuetos, 2012). Our study aims to compare the effect of word frequency and word orthographic neighbourhood density and its effects on eye movements in readers with dyslexia compared to typical readers (without dyslexia), specifically in the Czech language. Our results can reflect different strategies of visual word recognition in people with dyslexia and so it can tell us whether they use lexical and sublexical way in the same manner as typical readers and these are affected by some more generalized deficit (such as deficits in automatization and other related functions) or there is for example a specific deficit in one of these two paths that is not used properly.

2 Methodology

Eye movements during silent sentence reading were recorded during our study. Differences in gaze duration, first fixation duration and number of fixations of the target words are at the centre of our interest. 80 target words differing in word frequency (low - high) and word orthographic neighbourhood density (low - high) were included in the frame sentences created originally for the study and presented with the use of Tobii Pro Lab software. Data were collected with Tobii TX 300. All participants were Czech native speakers, university students or absolvents with no developmental reading disorder in the control group and with the developmental reading disorder in a clinical group. Apart from the eye tracking a set of tests focused on reading and related skills was administered individually with each participant. A linear model was used for the statistical analysis.

3 Results

The contribution describes the ongoing study, so the data collection has not been finished at the time of submission of the abstract. Therefore the results will be presented during a conference talk. Preliminary results

---

[1]More frequently used words in a language are recognized aster and with less effort than less frequently used words.

[2]Orthographic neighbourhood density is a word property that expresses how similar is the word to other words in a language (the more orthograghic neighbours the word has the more similar to other words in a language is). The effect of this property in literature is various and it is affected by many factors such as the type of task used for its evaluation.

based on the first part of collected data suggest a strong effect of dyslexia the most prominent in gaze duration and total fixation duration as well as the effect of word frequency similar in both clinical and control group. The orthographic neighbourhood density seems to have a weak or no effect on our variables.

4 Conclusion

It is necessary to deepen our knowledge about reading especially in people with dyslexia who struggle in using this skill in everyday life. Based on our research we can suggest whether there are some differences in using lexical and sublexical reading strategies in that group compared to typical readers. We hypothesize that people with dyslexia can use lexical and sublexical reading strategies in the similar manner as typical readers but their performance is affected by a more generalized deficit that affects both ways.

The final results will be discussed during the conference talk. It is worth noting that only high-functioning dyslexic people are involved in the study, so from the results, we can find out which deficits persist despite improving reading performance as well as which reading strategies seem to be effective enough for this group of people. We can use our knowledge in general dyslexia discourse as well as in creating/translating/evaluating test/diagnostics materials or adapted learning materials. Also, it can show us the main points on which it is necessary to target the intervention methods used in dyslexic people in the Czech language.

Literature

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1), 5–17. https://doi.org/10.1080/13603110600574322

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language*, *80*(3), 617–626. https://doi.org/10.1006/brln.2001.2637

Guitard, D., Miller, L. M., Neath, I., & Roodenrys, S. (2024). Set size and the orthographic/phonological neighbourhood size effect in serial recognition: The importance of randomization. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, *78*(1), 9–16. https://doi.org/10.1037/cep0000320

Jošt, J. (2009). *Oční pohyby, čtení a dyslexie*. Fortuna.

Laxon, V., Gallagher, A., & Masterson, J. (2002). The effects of familiarity, orthographic neighbourhood density, letter-length and graphemic complexity on children's reading accuracy. *British Journal of Psychology*, *93*(2), 269–287. https://doi.org/10.1348/000712602162580

Nemati, F., Westbury, C., Hollis, G., & Haghbin, H. (2022). The Persian Lexicon Project: Minimized orthographic neighbourhood effects in a dense language. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-022-09863-x

Suárez-Coalla, P., & Cuetos, F. (2012). Reading strategies in Spanish developmental dyslexics. *Annals of dyslexia*, *62*(2), 71–81. https://doi.org/10.1007/s11881-011-0064-y

Vyhnálek, M., Brzezný, R., & Jeřábek, J. (2006). Oční pohyby u specifických vývojových dyslexií. *Česká a Slovenská Psychiatrie*, *102*(5), 256–260.

Analyzing texts with graph theoretical and other innovative methods

Mária Markošová, Daniel Petriľák

Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University,
Bratislava, Mlynska dolina, 842 48, Slovakia
markosova@ii.fmph.uniba.sk

Abstract

We study various texts using less traditional methods, such as the ones borrowed from the graph theory. We combine them with traditional linguistic analysis. We created word webs from the written texts of healthy authors and author with aphasia to find, whether the signs of aphasia can be recognized. Last, but not least, we also used punctuation marks distribution analysis, introduced recently by Kulig et al.

1 Introduction

In recent years, there were several efforts to apply new methods of text analysis and combine them with the standard linguistic studies (Cancho et al., 2001; Dorogovtsev et al., 2001; Markošová, 2008). For example, methods from the graph theory. Here we created positional word webs Cancho et al. (2001) from several texts of healthy authors and one author with aphasia.

To create a word web, one considers words as nodes and if the words are neighbors, they are connected by an edge. The final word web is a binary complex network with a power law degree distribution, which can be found as a result of dynamical equation model solution of the network growing by preferential node attachment (Barabasi et al., 1999; Markošová, 2008).

In addition, several other averages and graph characteristics are usually measured and calculated. Graphlet analysis can be applied as well (Przulj, 2007). Graphlet is a small (up to five nodes) induced connected subgraph of a graph, generalizing concept of the node degree in word webs.

The main goal of these studies is to recognize signs of aphasia, a brain disease, which affects abilities of writing and communicating. It appears after brain injuries or brain tumor. There are many types of aphasia (Broca, Wernicke…), with different stress on different inabilities. We created word webs based on the three books (RMP, PFP, and SF) of an American author PW. He suffered a milder form of global aphasia. Here first two books were written before aphasia and the last one (SF) as a therapeutic effort to gain back some communication skills after aphasia. This appeared to be quite successful in his written language. To compare the stability of word web properties in time, we also studied several texts (U, CK, KZ, VT) of Slovak writer DD. DD has no aphasia.

Recently, Kulig and others introduced punctuation mark distribution analysis (Kulig et al., 2017). We used their method to find out differences in the texts before and after aphasia as well.

2 Results

Here we define word web like this: Unlike in classical positional word webs, in our web, word is a node, but different forms of the same word are taken as different words. This does not matter that much in English, but in Slovak, it does. The reason is that people with aphasia are prone to use various forms of words incorrectly. We do not care about punctuation marks in word web creation, the same way as in Cancho et al. (2001). We also calculated number of graphlets, and graphlet based measures called agreements (Przulj, 2007). Standard linguistic analysis of all texts has been provided as well. Our newest results concern punctuation mark distributions (Kulig et al., 2017). We measured how many words occur between two punctuation marks, either specified (such as coma for example) or not. Because people with aphasia use language differently, we suppose to find differences in these distributions too.

2.1 Word webs

It is not possible to include all big tabs here; therefore, we only mention some significant results. All texts of PW have from 4700 up to 5700 number of unique words. In the DD texts, it was from 5300 up to 6000 unique words. There are no significant differences in RMP, PFP, SF in clustering coefficient, graph density, average shortest path, which could be assigned to aphasia. The only significant differences are in network diameter, which is shortest in SF, the book written after aphasia occurred. The diameter difference between SF and RMP is almost 43 percent, and between SF and PFP almost 29 percent. We did not observe such differences in the texts of DD. The maximal diameter difference here is about

10 percent. Word web analysis of DD books show, in general, more similar graph parameters as in PW books. Numbers of 2, 3, 4 node graphlets measured in PW word webs also do not reflect differences due to aphasia. Graphlet based measures called geometric and arithmetic agreements were calculated between RMP and SF being 0.83 and 0.87, respectively. This indicates great similarity of the graphlet structure of both word webs. The similar values of agreements for PFP and SF also indicate graphlet structure similarity.

Linguistic statistics of the texts of PW also does not show differences due to aphasia. It is known that people with aphasia create short sentences, but surprisingly, the greatest average sentence length was measured in SF. It was almost 1.6 times longer than in PMP, and 1.26 times longer as in PFP. In DD books, the differences in average sentence length were not that big. The greatest difference is between U and KZ, the average sentence in U is 1.2 times longer as in KZ.

Last, but not least we studied distributions of punctuation marks. Here we found striking tendency in all cases, that these distributions for SF text are less steep than the other ones (Fig.1, Fig.2). All of these distributions follow line in lin- log scale, indicating rule

$$y = c \cdot exp\,(a{*}x) \ , \tag{1}$$

where c is a constant.

Distributions of "," and "." reflect the fact, that SF has higher average sentence length then RMP (Fig.2). These results are recent and we continue to study punctuation mark distributions in more details.
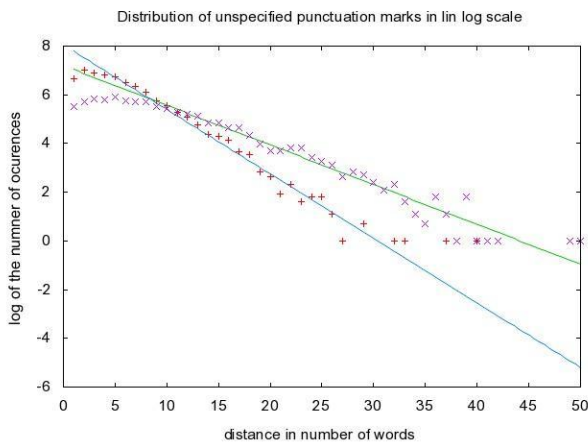
Fig. 1. Distribution of punctuation marks for RMP (+) and SF (x). RMP distribution follows function f(x)=0.26x+8.05, SF g(x)=-0.16 x+7.21.
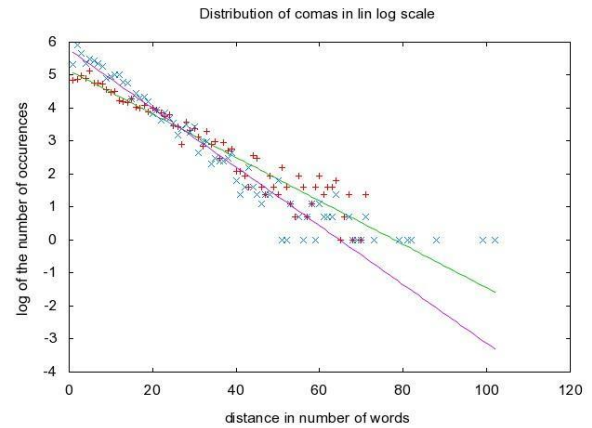
Fig. 2. Distributions of comas for RMP (+) and SF (x). RMP distribution follows function f(x)=-0.09x+5.77, SF g(x)=-006 x+5.14.

Discussion

Word web studies, together with linguistic studies of our data do not show persuasively, that our methods are able to find signs of aphasia in the text. There is due to several reasons:

1. *We do not have the raw data*. It is known, that another writer, the wife of PW, edited the SF book written after aphasia occurred. This could be a reason, that the average length of sentence measured in the number of words is, unexpectedly, the longest in the SF book. Maybe the editor simply put punctuation mark between sentences, which looked too short and made it longer.

2. *Binary word webs are too coarse models*. Therefore, we turned our attention to the weighted word webs. People with aphasia have tendency to repeat the same word sequence many times. This could be seen in the weighted models, which is our next step of the word web studies.

3. *Style of the book* reflected in the measured graph and linguistic properties outperforms the differences caused by aphasia. If there are many dialogs for example, this shortens the average length of the sentence.

4. *More and larger raw data are necessary*. However, unfortunately, such datasets are a problem, because sensitive data are under strict regulations. Moreover, people with aphasia have communication problems, and thus are

not able to create long texts. We gained additional datasets of raw unedited data of the spoken language. However, each dataset has only up to around 500-700 words. These datasets are also under scrutiny.

References

Barabasi, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512. https://doi.org/10.1126/science.286.5439.509

Dorogovtsev, S. & Mendes, J. F. (2001). Language as an evolving word web. *Proc. Royal Soc. B, 268*, 2603-2606. https://doi.org/10.1098/rspb.2021.1824

Ferre-i-Cancho, R. & Sole, R. (2001). The small world of human language. *Proc. Royal Soc.*, *268*, 2261-2265. https://doi.org/10.1098/rspb.2001.1800

Kulig, A. et al. (2017). In narrative texts punctuation marks obey the same statistics as words. *Information Sciences*, *375*, 98-113. https://doi.org/10.1016/j.ins.2016.09.051

Markošová, M. (2008). Network model of human language. *Physica A, 387*, 661-666. https://doi.org/10.1016/j.physa.2007.09.027

Przulj, N. (2007). Biological network comparison using graphlet degree. *Bioinformatics*, *23*, 177. https://doi.org/10.1093/bioinformatics/btl301

Using Eye-Tracking Analysis on Secondary School Students for mapping the correlation between their better physics grades and an ability to solve graphs

Jakub Hubáček[1], Libor Machala[1], Jana Škrabánková[2]

[1]Department of Experimental Physics, Faculty of Science, Palacky University
17. listopadu 1192/12, 779 00, Olomouc, Czech Republic
[2]Department of Physics, Faculty of Science, University of Ostrava
30. dubna 22, 702 00, Ostrava, Czech Republic
Jakubhubacek2@gmail.com, Libor.machala@upol.cz, Jana.skrabankova@osu.cz

Abstract

Eye-tracking technology is used not only in marketing or economics, but it is applied in other areas. We integrated the Eye-tracking technology to pedagogy-didactic research. In the study, we applied Eye-tracking techniques to investigate the graph-reading abilities of secondary school students in selected physics tasks and we were able to map the graph-reading abilities of students with better physics grades. We formulated two hypotheses. Students assessed with an "excellent" score solved tasks more successfully than those with only a "good" assessment. These students required a longer time to solve the tasks on average. We also discussed gender differences. There were no significant differences between boys and girls in the analysis. The discussion contains the results of these analyses and involves recommendations for the inclusion of graphs in the physics education. Developing the skill and knowledge to work with and interpret graphs is an especially important aspect of education in physics and contributes to the student's development of thinking about physics.

1 Introductory chapter

The skill of reading the content of graphs is a very important component in science education. As teaching is driven by modern trends, graphs and their inclusion in the process have great relevance, especially in mathematics, physics, but also in other engineering disciplines.

Several studies in the past have investigated students' problem-solving strategies and abilities to read physics graphs, their success rates, or comparisons between genders.

An earlier eye-tracking study by one of the authors reported that students find reading graphs in physics problematic and are thus less successful in solving physics problems (Škrabánková et al., 2020). Some other studies have also shown the difficulty of understanding the physics concepts contained in graphs. In some, the authors report that it is more natural for students to read graphs in mathematical concepts rather than in physical ones (Planicic et al., 2012). Others, in turn, show that there are minimal differences in the success rate in reading graphs between physics students and students in other disciplines such as psychology, although this should not be the case (Susac et al., 2018).

Teachers themselves need to know how to present and interpret graphs to students for maximum understanding. Another study states that it is important to maintain a top-down distribution (Madsen et al., 2012).

2 Hypotheses

Based on our assumptions and a search of previous studies, we formulated two hypotheses.
Hypothesis 1: Students who are more successful in physics will be more successful in problem solving in graphs.
Hypothesis 2: Students who are more successful in physics will not need more time to solve problems in graphs.

The hypotheses were formed based on previous research by one of the authors or his colleagues and information we had after studying articles on this topic. The hypotheses were then tested using Eye-tracking analysis, statistical analysis, and scientific research.

3 Methodology

We conducted Eye-tracking measurements with 30 students from different secondary schools (grammar schools) aged 16 to 18. Data collection took place during a covid pause when the students were home-schooled. Students were presented with 10 problems containing physics graphs with a choice of four answers and one correct answer. A Gazepoint GP3 instrument was used for data collection. For Eye-tracking analysis, the data were converted into OGAMA software.

Together with the results of previous studies, the data were interpreted and discussed.

3.1 Eye-tracking analysis

After data conversion, the Eye-tracking analysis contained 3 important elements. These were Attention Maps, Areas of Interest and Sequence Chart.

3.2 Statistical analysis

Statistical analysis included the unpaired two-samples Wilcoxon test and the Kruskal-Wallis test, at a statistical significance level of 95 %. Similarities and statistical significance between student types were assessed in this analysis. Based on both analyses, we were able to discuss the confirmation or denial of the hypotheses.

4 Results

Hypothesis 1: Students who are more successful in physics will be more successful in problem solving in graphs was confirmed.
Excellent score students were on average 30% more successful in problem solving.
Hypothesis 2: Students who are more successful in physics will not need more time to solve problems in graphs was disproved. Excellent score students took longer to solve problems than B-grade and C-grade students.
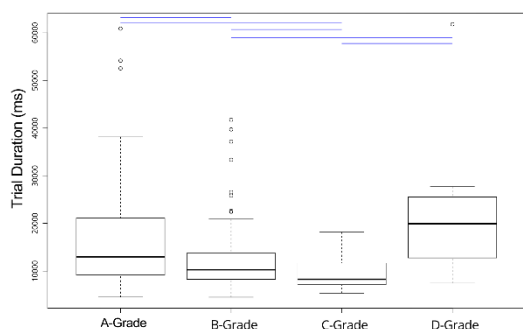
Fig. 1 Boxplot of time needed to solve tasks according to the student's previous grades in physics.

The measurement results can be seen in Figure 1 (Time duration among students by grade) and Figure 2 (Correct answers among students by grade).
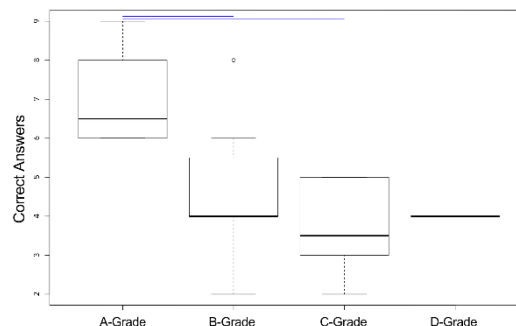
Fig. 2. Boxplot of number of correct answers according to the students' previous grades in physics

5 Discussion

Graphs have an irreplaceable role in physics education. If students are to survive in the modern job market, they must adapt to technological trends that emphasize technical and science education. Physics graphs belong to these conceptions.

We compared the differences in solving problems containing physical graphs among students in relation to their grade level in physics. Conclusions were drawn based on Eye-tracking and statistical analysis.

We rank the following two findings as the most interesting. The first is the fact that students who were more successful in solving the problems in graphs took more time to solve the problems.

The second finding is that the data did not show any significant differences between girls and boys in terms of time spent solving problems in graphs or success rates. Girls, however, were more likely to return to the assignments for each task. This finding may be linked to the fact that girls are generally more careful and simply strive for better grades than boys. However, it is contradicted by some sources which report that boys are more interested in physics and are generally more interested in physics.

6 Conclusion

In this study, we used Eye-tracking Analysis to investigate the ability of general secondary school students to solve problems involving physics graphs. We compared data from several studies with our results and discussed, in a broader context, the differences among students in their strategies and abilities to solve problems involving physical graphs. Eye-tracking analysis was used to interpret the results.

Literature

Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics-Physics Education Research*, 8(1), 010122.

Planinic, M., Milin-Sipus, Z., Katic, H., Susac, A., & Ivanjek, L. (2012). Comparison of student understanding of line graph slope in physics and mathematics. *International journal of science and mathematics education*, *10*(6), 1393-1414.

Skrabankova, J., Popelka, S., & Beitlova, M. (2020). Students' Ability to Work with Graphs in Physics Studies Related to Three Typical Student Groups. *Journal of Baltic Science Education*, *19*(2), 298-316.

Susac, A., Bubic, A., Kazotti, E., Planinic, M., & Palmovic, M. (2018). Student understanding of graph slope and area under a graph: A comparison of physics and nonphysics students. *Physical Review Physics Education Research*, *14*(2), 020109.

Using Mobile Eye Tracker in Spatial Orientation Research

Kateřina Jochecová

Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00 Brno, Czech Republic
katerina.jochecova@mail.muni.cz

Extended Abstract

This abstract is focused on proof of concept for using eye trackers for wayfinding research. Here are defined problems that will be discussed at the presentation.

With mobile eye trackers becoming widely available, cognitive scientists are exploring the possibilities of getting more hard data in the world outside of the laboratories, enhancing the ecological validity of their research. These eye trackers are easy to use and practical to wear, as they are almost wireless. Their sampling frequency is incomparable to the remote eye trackers, but for the types of tasks that are usually executed outside, it's sufficient. Nevertheless, some argue that the attractivity of portable eye trackers might be overrated (Hessels et al., 2020).

Lately, it became obvious that there are many challenges regarding outdoor eye tracking data collection and analysis (Evans et al., 2019). The most important question is how precise the measurement in the terrain really is. The volatile light conditions (Binaee et al., 2021), bumps and shakes of the glasses (Niehorster et al., 2020) and change of the viewing angle through the walk need to be taken into consideration during experimental preparation. With the extensive amount of manual work needed for data analysis it might seem that in the end the work behind correct use of mobile eye trackers might not be worth it. But there is a modern solution that can at least partly help by transferring research to the virtual world.

Research shows that from cognitive point of view iVR (immersive virtual reality) is a suitable substitution for the real world in the context of spatial orientation research (Dong et al., 2021; Drewes et al., 2021; Stachoň et al., 2022), which might open a whole new approach to measuring eye tracking data during wayfinding. Unfortunately, the available options of eye tracking measurement in iVRs poses a completely new challenge for data collection and analysis, as most of the commercially available HMDs (head mounted displays) with integrated ET are not suitable for serious research. Knowing exactly what kind of eye behaviour is happening is crucial for correct data interpretation, but most newer HMDs have integrated self-made ET, which works as an absolute black box to users and furthermore have no possibility to export any of the ET data. Even the calibration accuracy itself is debatable (Adhanom et al., 2023). These HMDs are thus hardly if ever usable for research. Some of the commercial headsets (e.g., Pico, PS) have ET provided by Tobii, who lets you, for a generous fee, export raw data apart from the basic heatmap available. Currently many scientists are trying to program their own eye trackers.

Acknowledgement

2. Literature

Adhanom, I.B., MacNeilage, P. & Folmer, E. (2023). Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. *Virtual Reality, 27*. https://doi.org/10.1007/s10055-022-00738-z

Binaee, K., Sinnott, C., Capurro, K.J., MacNeilage, P. & Lescroart, M.D. (2021). Pupil Tracking Under Direct Sunlight. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 18, 1–4. https://doi.org/10.1145/3450341.3458490

Dong, W., Qin, T., Yang, T., Liao, H., Liu, B., Meng, L., Liu, Y. (2021). Wayfinding Behavior and Spatial Knowledge Acquisition: Are They the Same in Virtual Reality and in Real-World Environments? *Ann. Am. Assoc. Geogr.* https://doi.org/10.1080/24694452.2021.1894088

Drewes, J., Feder, S., & Einhäuser, W. (2021). Gaze During Locomotion in Virtual Reality and the Real World. *Front. Neurosci, 15*. https://doi.org/10.3389/fnins.2021.656913

Hessels, R. S., Niehorster, D. C., Holleman, G. A., Benjamins, J. S., & Hooge, I. T. C. (2020). Wearable Technology for "Real-World Research": Realistic or Not? *Perception, 49*(6), 611-615. https://doi.org/10.1177/0301006620928324

Niehorster, D.C., Santini, T., Hessels, R.S., Hooge, I. T.C., Kasneci, E. & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods, 52*. https://doi.org/10.3758/s13428-019-01307-0

Stachoň, Z., Jochecová, K., Kvarda, O.,..Šašinka, Č. (2022). *The Possibilities of Using Virtual Environments in Research on Wayfinding*, Research Square. https://doi.org/10.21203/rs.3.rs-2045842/v1.

Cross Cultural Research in Cartography

Stachoň Zdeněk[1], Šašinka Čeněk[2], Popelka Stanislav[3], Jiří Čeněk[2], Lacko David[2]

[1]Faculty of Science, Masarykova univerzita, Kotlářská 2, 602 00, Brno, Czech Republic
[2]Faculty of Arts, Masarykova univerzita, Arna Nováka 1, 602 00, Brno, Czech Republic
[3]Dept. of Geoinformatics ,Univerzita Palackého v Olomouci, 17. listopadu 50, 779 00, Olomouc, Czech Republic
zstachon@geogr.muni.cz, ceneksasinka@mail.muni.cz, stanislav.popelka@upol.cz,
144946@mail.muni.cz, david.lacko@mail.muni.cz

Abstract

This contribution is focused on research on cross-cultural differences in the process of visual search on maps. It concentrates mainly on searching for point features on maps using eye-tracking methodology, offering a perspective on the potential applications of eye-tracking data in theoretical cartographic research. The stimulus material consists of two unique sets of symbols and color schemas designed to highlight differences in their visual searchability. These proposed symbol sets vary in the parameters of different visual variables (e.g., shape, color, etc.) and simulate symbol sets for large-scale maps. The findings, in general, support previous studies, such as the significant effect of color contrast between the target symbol and the topographic background, the considerable attraction to the red color, and the impact of the relative position of the target symbol on the speed and success of its identification. A detailed examination of the observed effects was made possible by objective data obtained through the eye-tracking method. The results have the potential to contribute to a deeper understanding of the visual search process in map reading and reveal the influence of selected visual variables and the relative position of the target on strategies used for visual symbol search. Future research could focus on other variables affecting visual search. For instance, the influence of various systems of written language representation might introduce distinct visual search strategies or varying responses to the visual variables used on maps.

Acknowledgements

VR for Experimental Psychology:
A User-Friendly Solution for conducting eyetracking and pupillometry study

Beata Pacula-Leśniak, Agata Szymańska, Laurent Beaupoil, Magdalena Igras-Cybulska, Artur Cybulski

Emotion and Perception Lab, Jagiellonian University, Gołębia 24, 31-007, Krakow, Poland
AGH University of Krakow, al. Adama Mickiewicza 30, 30-059, Krakow, Poland
Interactive Crafts sp. z o. o, Lublin, Poland
beata.pacula-lesniak@uj.edu.pl, agata.szymanska@doctoral.uj.edu.pl, laurent.beaupoil@uj.edu.pl, migras@agh.edu.pl,
a.s.cybulski@gmail.com

Abstract

We introduce the Free-viewing VR Presentation Tool, developed in Unity – the user -friendly solution designed to facilitate the creation and execution of VR experiments. Additionally, we share initial findings from a free-viewing eye tracking experiment conducted in a virtual environment utilizing this tool.

1 Introduction

Virtual reality (VR) devices provide an exceptional platform for presenting stimuli, delivering immersive experiences and unparalleled control over the human visual field. Despite these advantages, the utilization of VR in experimental psychology is limited due to the lack of support for VR headsets in commonly used stimuli presentation software like Presentation, Eprime, or PsychoPy (see eg. Bebko & Troje, 2020; Brooke et al., 2020; Deuchler et al, 2023; Peirce, 2019; Watson et al., 2019). Our goal was to create a solution that allows users without prior experience in programming in virtual environments to effortlessly build stimuli presentation protocols. In the current work we present the user-friendly Free-viewing VR Presentation Tool, built in Unity, aimed at facilitating VR experiment creation and execution; as well as the preliminary results of the free-viewing eye tracking experiment adapted to a virtual environment with the usage of the tool.

2 The Free-viewing VR Presentation Tool

Our software supports the creation and implementation of test routines, data acquisition, recording, and storage. In particular, it can present stimuli for a predefined time, log user responses, and send triggers to external measurement devices like EEG or physiological measurement equipment, which is needed to synchronize the data from different channels. Experiments in the VR environment could be designed and conducted by specifying content and time course

with a collection of one-line commands assembled in a configuration file in .csv format, that is loaded and initialized using Inspector GUI - the front end of the Unity pre-defined scenario.

The data acquisition is made with the use of Pupil Labs HMD Eyes Unity package components managed by custom scripts developed for the project and with concurrent use of Pupil Labs Pupil Capture software and Pupil Labs HTC Vive Pro Binocular Add-on. Before the research session, the stimuli sequence scenario gets loaded within Unity to the prepared scriptable object from researcher provided .csv via Inspector GUI of the Scenario Manager component that allows for convenient session necessities setup. Each scenario comprises record rows containing researcher-predefined functional trigger IDs for managing eye-tracking recording and calibration and unique stimuli trigger IDs with their image or video source paths, duration, scale, and background color. During the research session parts that have recording turned on the software sends the just-loaded trigger ID to the specified IPv4 address at the specified port, as well as to the Pupil Capture in form of an annotation. After each recording, the stored eye-tracker data are ready for loading in Pupil Labs Pupil Player software and from there it is possible to export the data of the desired researcher-predefined parameter types into the human-readable .txt files.

The tool supports the most common picture (.jpg, .png) and video (.asf, .avi, .mov, .mp4) formats.

From the participant point of view, the stimuli are shown in 2D on a 'virtual screen'. There is no visible avatar of the participant and moving around the virtual space is not possible - the only way of interaction is stopping/continuing the presentation of the stimuli or giving a simple choice answer. However restricted, those possibilities correspond almost perfectly with the non-virtual eye-tracking experiment setups and in this way give additional value, allowing to easily replicate the already existing experiment flows in a controlled environment, and then further adapt. During the experiment, the experimenter has the preview of the eye-tracking data in a form of the participant gaze presented

in real time on the image preview. The accuracy of the gaze mapping depends on calibration precision and allows to additionally evaluate the data quality. The experimenter has also preview of the log data with the information of the scenario loading, stimuli loading, and any potential errors in the configuration file.

3 Research

We have utilized the set-up in an eye-tracking study, where 100 participants (50 men and 50 women, aged 18-30) viewed a subset of 120 pictures from emotional pictures databases - IAPS (Lang et al., 2008), NAPS (Marchewka et al., 2014), GAPED (Dan-Glauser & Scherer, 2011), EmoPics (Wessa et al., 2010). We manipulated the visual saliency of the key objects with the MATLAB Shine toolbox to obtain two conditions: saliency increased or unchanged, and each person saw only one saliency version of each picture. Additionally, they watched 75 emotional videos (Ack Baraly et al, 2020) and listened to 80 emotional sounds (Bradley & Lang, 2007). In the initial analysis, we focused on the measurements of pupil diameter changes (pupillometry). Pupillometry grants access to a fast-growing, non-invasive measure of cognitive changes. Indeed, if controlling the light condition, there are still changes in size associated with a plethora of arousal-linked cognitive functions, such as emotions, surprise, or decision-making (Zénon, 2019). Certainly, both pupil dilation and gaze patterns can easily be measured together, even in VR, using the same infrared camera. In case of the described study, we were using fixations predefined in Pupil Labs that have been further processed in MATLAB. It is also possible to define AOIs using the solution offered by Pupil Capture software or post-factum, externally (in our case, in MATLAB).

Our results confirm that the saliency manipulation has a significant impact on pupil dilation in the VR setup; still, the impact of the emotional load is visible - the pupil becomes more dilated in case of negative stimuli than neutral and positive. There is no interaction between low-level features and the impact of emotions. Importantly, the results obtained in the VR environment are perfectly in line with the results from the free-viewing experiment with the use of a stationary eye-tracker (EyeLink Portable Duo), indicating that the traditional eye-tracking methodology could be adapted to virtual environments using the Free-viewing VR Presentation Tool. Moreover, the results confirm that the intensity of emotional arousal in response to a specific stimulus can be directly read from changes in pupil size during the presentation in the virtual environment.

Literature

Ack Baraly, K. T., Muyingo, L., Beaudoin, C., Karami, S., Langevin, M., & Davidson, P. S. (2020). Database of emotional videos from Ottawa (DEVO). *Collabra: Psychology*, *6*(1), 10.

Bebko, A. O., & Troje, N. F. (2020). bmlTUX: Design and control of experiments in virtual reality and beyond. *i-Perception*, *11*(4), 2041669520938400.

Bradley, M. M., & Lang, P. J. (2007). *The International Affective Digitized Sounds: Affective ratings of sounds and instruction manual* (Technical Report No. B-3). Gainesville, FL: University of Florida, NIMH Center for the Study of Emotion and Attention.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior research methods*, *52*, 455-463.

Dan-Glauser, E. S., & Scherer, K. R. (2011). *Behavior research methods*, *43*(2), 468-477.

Deuchler, J., Hettmann, W., Hepperle, D., & Wölfel, M. (2023, March). Streamlining physiological observations in immersive virtual reality studies with the virtual reality scientific toolkit. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 485-488). IEEE.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *NIMH Center for the Study of Emotion and Attention*, *1*, 39-58.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). *Behavior research methods*, *46*(2), 596-610.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13.

Watson, M. R., Voloh, B., Thomas, C., Hasan, A., & Womelsdorf, T. (2019). USE: An integrative suite for temporally-precise psychophysical experiments in virtual environments for human, nonhuman, and artificially intelligent agents. *Journal of neuroscience methods*, *326*, 108374.

Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfelder, S. (2010).. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *39*(Suppl. 1/11), 77.

Zénon, A. (2019). Eye pupil signals information gain. *Proceedings of the Royal Society B*, *286*(1911), 20191593.

Democratizing Eye-Tracking Data Visualization with GazePlotter

Michaela Vojtěchovská, Stanislav Popelka

Department of Geoinformatics, Palacký University Olomouc
17. listopadu 50, 771 46, Olomouc, Czech Republic
mail@vojtechovska.com, stanislav.popelka@upol.cz

Abstract

Eye-tracking technology provides invaluable insights into cognitive processes, preferences, and mental states by analyzing individuals' gaze patterns as they explore various stimuli. This capability is pivotal across various fields, from neuroscience research and educational methodologies to experimental psychology. However, the accessibility of eye-tracking analysis has been limited by the need for proprietary software or advanced programming skills, hindering research progress by excluding non-technical researchers from unlocking the full potential of eye-tracking data. To bridge this gap, we introduce GazePlotter, an innovative tool designed to democratize Areas of Interest (AOI) analysis through automated and user-friendly visualization techniques, particularly scarf plots. Scarf plots, or sequence charts, graphically represent the sequence of gazes over time, facilitating the comparison of gaze patterns across participants. By leveraging GazePlotter, we have made it easier to analyze dynamic AOIs—previously a challenging task—enhancing studies involving interactive interfaces or video stimuli. GazePlotter does not require preliminary data transformation or technical expertise, offering automatic data recognition and transformation. This accessibility empowers researchers from diverse backgrounds to engage with eye-tracking data, enhancing the inclusivity and scope of research findings. Key features of GazePlotter include support for dynamic AOIs, participant grouping, customization options, and the ability to integrate with other AOI analysis tools. Furthermore, as an open-source platform, GazePlotter invites contributions from the community, ensuring its continuous improvement and adaptability to user needs. This presentation will highlight GazePlotter's interface, demonstrate its capabilities through case studies, and discuss future development directions driven by community feedback. By simplifying the process of eye-tracking analysis, GazePlotter aims to unlock new perspectives and foster a more inclusive research environment, ultimately advancing our understanding of visual cognition.

Exploring Map Reading Strategies Across Various Groups

Markéta Beitlová

Department of Geoinformatics, Palacký University Olomouc
17. listopadu 50, 77146, Olomouc, Czech Republic
marketa.beitlova@upol.cz

Abstract

The research aimed to generally assess the effectiveness of cartographic communication models. We analyzed how different individuals engage with maps, including those who create them and those with varying levels of cartographic knowledge. The participants included a broad range of individuals, some of whom had experience in designing maps and others who were less familiar with cartographic principles. The methodology involved observing how participants interact with maps under different conditions, including open-ended exploration and completing specific tasks. The outcomes suggested that those who had created the maps were more adept at navigating and interpreting them correctly compared to other participants. Interestingly, individuals with less experience in cartography tended to require more effort to understand the maps, as indicated by their viewing patterns, yet they were more attentive during uninstructed viewing sessions. A detailed examination of how all participants focused on various map features showed a more structured approach by those with cartographic training. A key part of our study was analyzing how similar the map reading strategies were across all participants, utilizing advanced analytical tools to examine the order in which they focused on different map areas. This comparison provided insights into the commonalities and differences in map engagement strategies. Our findings offer a broad overview of how people from different backgrounds interact with maps, reflecting a wide range of engagement levels with cartographic content. Although there was a notable alignment in how maps were read across the board, distinct individual strategies emerged, suggesting a diversity in understanding and maps. This broad examination sheds light on the variability in cartographic comprehension and opens up discussions on enhancing cartographic communication models.

Acknowledgements

AI that looks, but does not see

Martin Takáč[1,4], Alistair Knott[2,4], Mark Sagar[3,4]

[1]Comenius University in Bratislava, [2]Victoria University of Wellington, [3]University of Auckland
[4]Soul Machines Ltd., Auckland, New Zealand
martin.takac@soulmachines,com, alistair.knott@soulmachines.com, mark.sagar@soulmachines.com

Abstract

Recently we have seen proliferation of very realistic DeepFake videos based on GenerativeAI models trained on videos, i.e. learning probabilities of pixel transitions. In this contribution, I am comparing these models to an interactive AI with a biologically inspired cognitive architecture built for real interactions – BabyX – a virtual infant simulator. This simulator contains a rich physiological model interacting with the cognitive architecture via modulatory circuits. As opposed to DeepFakes, the visual rendering results from simulated cognitive operations and physical movements, exemplified in the paper with the visual processing pathway.

1 Introduction

Recently OpenAI released Sora[1] – a system for generating minute-long high fidelity videos from text instructions. Sora is based on a pre-trained diffusion transformer that starts with visual noise and iteratively denoises the image, conditioned on a supplied text prompt. The process is made more effective by using spacetime latent patches as building blocks (Liu et al., 2024). Other video generation systems vary in the underlying technique (for review, see Karaarslan & Aydi, 2024), but all of them are trained directly on videos, and learn to generate videos as output. They essentially learn to render and transform pixels. Hidden factors causing the pixels move in certain ways are not represented directly but are inferred by powerful learning mechanisms during training. For example, eye movements of the characters do not correspond to actual deployments of attention: they play no functional role in a visual process. In short, they seemingly look, but don't see anything.

In addition, impressive as they are, these videos only support unidirectional flow of information – the audience is simply watching the movie without an ability to steer its content in real time. More interesting in this sense are animated characters (avatars) able to interact with humans in real time, whether in computer games or for more practical purposes, such as customer support, e-commerce or entertainment. In this paper, I present a different, embodied, approach to building realistic visual interactions – BabyX, a hyperrealistic simulator of a human baby designed for interactions with humans. The overall experimental framework is explained in Section 2. Section 3 provides a brief description of the cognitive architecture, emphasizing its embodied nature. Section 4 focuses on the visual pathway.

2 The BabyX experimental framework

BabyX is a hyperrealistic virtual simulation (animated 3D model) of an 18-24-month-old baby. It is a research platform for studying development of cognition and early cooperation between a child and its caregivers (Sagar et al., 2023a). Due to its modular architecture, it is also an experimental testbed for development of a human-like AI.

BabyX has a simulated brain and body, which is situated in a simulated 3-D environment. Both body and environment are rendered using computer animation techniques. An important part of the environment is a semitransparent touchscreen in front of the baby, which she can see and act on. This constitutes a shared space, in which BabyX and a human participant can interact e.g. by playing various games (see Figure 1). BabyX can reach to the board in her simulated world and move objects on the board with her finger. The human can move screen objects by clicking and dragging with a mouse.[2] BabyX can also see and hear the human user through a camera and a microphone.

---

[1] https://openai.com/sora
[2] An interested reader can find video demos at:
https://www.soulmachines.com/babyx_computational_linguistics/

3 Cognitive architecture

BabyX's cognitive architecture is a network of interconnected neural network modules, dynamical systems and state machines operating in parallel. These modules control a graphical simulation of a baby's phys-

Fig. 1. BabyX playing with sea animals in a shared interaction space.

ical body, including face, torso, limbs, and autonomic functions (breathing, heartbeat). BabyX has sensory systems (vision, hearing, touch, proprioception, interoception), coupled with models of cognitive processing (e.g. visual attention, visual object classification, category learning) and motor function (e.g. reaching-to-grasp, visual object tracking). She also has a model of emotions, that incorporates models of subcortical and cortical pathways, neurochemicals and emotional behaviours (for details, see Knott et al., 2022). And she has a model of memory, including working memory (WM), episodic long-term memory (LTM), and affective memory. Building on these, BabyX has (age-appropriate) language abilities: production, interpretation and learning. The system is specially tailored for real-time interactions.

4 Visual perception

BabyX segments her experience of the world into a stream of discrete "events", just as humans do (Knott & Takac, 2020). The event perception process is described in detail in Sagar et al. (2023b). Here we focus on a process that includes perception of both static scenes and moving objects.

BabyX's visual system consists of peripheral and focal vision. Peripheral vision includes a lower resolution image of the whole scene, while focal vision is a higher-resolution image of a foveated region of interest. The peripheral image is an input to the attention system that consists of a collection of saliency maps: some are based on salient features of the environment, others on movement, yet others express top-down salience (e.g. remembered location of objects of interest). All maps are combined, and the most salient region is selected for focal attention. Eye gaze is attracted to that location and the foveated region of interest is sent for downstream processing, e.g. to object classification system. The attention system includes

habituation, i.e. the saliency of the foveated region is temporarily suppressed after some time, which generates iterative saccadic eye movements. In this sense, unlike in Deepfakes, the eye movements of BabyX correspond to meaningful cognitive processing. Her other bodily movements are similarly driven by meaningful, realistic cognitive processing.

5 Conclusion

Recent advances in generative AI suggest that soon there will be many visual simulations of humans; some of them based on superficial pixel transitions, other building upon biologically inspired cognitive architecture. We hypothesize that systems whose behaviour is produced by underlying embodied models of human cognition and physiology will ultimately provide a better model of humans than the current generation of deepfake-type systems, and produce more realistic behaviour and interactions. No AI systems have yet truly escaped the 'uncanny valley'---but we believe systems whose behaviours are generated from realistic simulations have the best prospect of doing so. Whether this is true, needs to be tested empirically.

Acknowledgements

Literature

Karaarslan, E., & Aydin, O. (2024). Generate Impressive Videos with Text Instructions: A Review of OpenAI Sora, Stable Diffusion, Lumiere and Comparable Models. 10.36227/techrxiv.170862194.43871446/v1

Knott, A., Sagar, M., & Takac, M. (2022). The ethics of interaction with neurorobotic agents: a case study with BabyX. *AI and Ethics, 2*, 115-128.

Knott, A., & Takac, M. (2020). Roles for Event Representations in Sensorimotor Experience, Memory Formation, and Language Processing. *Topics in Cognitive Science, 13*(20), 187-205.

Liu, Y., Zhang, K., Li, Y., Yan, Z., Gao, Ch., Chen, R. Yuan, Z., Huang, Y., Sun, H., Gao, J., He, L., & Sun, L. (2024). Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models. arXiv:2402.17177v2 [cs.CV]

Sagar. M., Henderson, A. M., Takac, M., Morrison, S., Knott A., Moser A., Yeh W.-T., Pages N., & Jawed K. (2023a). Deconstructing and reconstructing turn-taking in caregiver-infant interactions: A platform for

embodied models of early cooperation. *Journal of the Royal Society of New Zealand, 53*(1), 148-168.

Sagar, M., Moser, A., Henderson, A. M., Morrison, S., Pages, N., Nejati, A., Yeh, W.-T., Conder, J., Knott, A., Jawed, K., & Takac, M. (2023b). A platform for holistic embodied models of infant cognition, and its use in a model of event processing. *IEEE Transactions on Cognitive and Developmental Systems, 15*(4), 1916-1927.

Explainability of Vision Transformer with Top-Down Connection

Štefan Pócoš, Iveta Bečková and Igor Farkaš

Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava
Mlynská dolina F1, 842 48 Bratislava, Slovak Republic
{stefan.pocos, iveta.beckova, igor.farkas}@fmph.uniba.sk

Abstract

We investigate the properties of our previously published model, RecViT, a vision transformer enhanced with top-down information flow. Since vision transformers have a higher level of inherent explainability, there are methods to examine the activations during image classification. Our main contribution is the comparison of the network activations given an adversarial example to the activations of the original image. Finally, we numerically evaluate the quality of attention maps on original and adversarial images using annotated segmentation masks.

1 Adversarial examples

Szegedy et al. (2014) were the first to demonstrate that modern neural networks can be fooled with adversarial examples (AEs) — images closely resembling clean data but containing a slight, carefully crafted noise, causing intentional misclassification. Many papers have since been dedicated to preventing the networks from being fooled by such inputs, yet none of them is able to guarantee complete robustness. This hinders the deployment of machine learning in safety-critical applications.

2 Recurrent vision transformer

To mitigate the effect that AEs have on the network classification, earlier, we proposed a recurrent vision transformer architecture (RecViT). The core idea of RecViT is to incorporate top-down information flow alongside the standard bottom-up approach. In practice, this means that after the initial prediction is made, the network "reconsiders" its decision, and in the subsequent iterations, it focuses on more specific, class-relevant features. After several iterations, we get the final output. As the RecViT is a vision transformer model, it inherits the attention mechanism and its properties. The attention heads in the network provide a simple means of extracting visual clues about what happens in the network during inference (Kashefi et al., 2023).

3 Comparison of saliency

Besides heightened robustness, using RecViT, we examine the activation space of the images. When plotting the attention weights for the input images in the form of heatmaps (see Fig. 1), we observe differences between AEs and the clean data. Usually, when dealing with the AEs, the network focuses more on seemingly irrelevant image regions.

Fig. 1: The attention map for the original vs the adversarial image.

Visual plots can be informative. However, they do not provide a means of making robust conclusions. Therefore, we use the segmentation masks provided in the Oxford-IIIT Pet Dataset (Parkhi et al., 2012) to numerically evaluate the differences between the activations of clean vs corrupted data.

Thus, after flattening the segmentation mask s and the attention scores a for a particular input image, we can simply calculate, for example, their cosine similarity:

$$\mathrm{sim}(\boldsymbol{s}, \boldsymbol{a}) = \frac{\sum\limits_{i=1}^{n} s_i a_i}{\sqrt{\sum\limits_{i=1}^{n} s_i^2} \; \sqrt{\sum\limits_{i=1}^{n} a_i^2}}. \tag{1}$$

4 Results

Our analysis of adversarial examples in the RecViT architecture focuses on their distinction from clean data inside the network. We visually assessed the differences between these data via extracting the attention weights in the self-attention mechanism, as well as we numerically confirmed our findings that the AEs do indeed evoke different activations in the RecViT. Further research could focus on finding the specificities in the RecViT and compare it to ViT.

Acknowledgement

References

Kashefi, R., Barekatain, L., Sabokrou, M., and Aghaeipoor, F. (2023). Explainability of vision transformers: A comprehensive review and new perspectives. arXiv:2311.06786 [cs.CV].

Parkhi, O. M., Vedaldi, A., Zisserman, A., and Jawahar, C. V. (2012). Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. In *International Conference on Learning Representations*.

Biologically motivated neural network UBAL in cognitive robotics

Kristína Malinovská and Jakub Mišovský

Centre for Cognitie Science, DAI FMPI,
Comenius University in Bratislava, Slovakia
Mlynská dolina, 84248 Bratislava
Email: kristina.malinovska@fmph.uniba.sk

The novel neural network model UBAL represents a more biologically plausible alternative to error-backpropagation. Since UBAL is a heteroassociator with a particular mechanism of activation propagation including self loops, it demonstrates interesting emergent phenomena such as the ability to generate meaningful reconstructions of the learned patterns. Cognitive robotics uses the paradigm of understanding by building and in this spirit explores and models the neural correlates of cognitive capacities in humanoid robots. In our modeling we tackled the understanding of motor actions and associations between different modalities such as vision, touch, and proprioception. In this paper, we outline the history of our modeling efforts as well as new prospects for our cognitive robotic models.

1 Introduction

Cognitive robotics aims at studying cognition of humans via "understanding by building" (Pfeifer & Scheier, 1999) employing connectionist bio-inspired models in robots that interact with and learn from complex and multimodal environments. It is well known the cognitive functions in the brain are built up in a hierarchical manner, the brain areas involved are processing information on lower levels such as low-level feature detection or motor control and pass it on to higher, so called association areas. The common coding theory (Hommel et al., 2001) suggests that there is a common representational base for perception and action (motor performance). The perception of action automatically activates its motor component and vice versa. The common coding framework might also be considered the means for sensorimotor simulation (Barsalou, 1999). A phenomenon closely related to common coding is the so called mirror neuron theory (Rizzolatti & Sinigaglia, 2010) which claims our brain has special mechanisms that allow us to resonate with or step into the shoes of another person thus providing us with deeper understanding of the observed actions.

Modeling of cognitive capacities in humanoid robots within the embodied paradigm of cognition builds up on robots own senses, competences and experience. In our research we have focused on building and connecting these capacities in a modular way, utiliz-
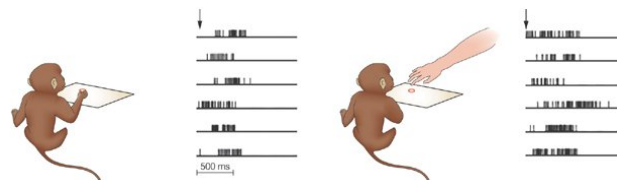
Fig. 1: Illustration of macaque monkey's mirror neuron responses - motor neurons are firing due to visual stimulation (Rizzolatti et al., 2009).

ing different neural network architectures and learning paradigms. This is something in contrast with current trend of end-to-end deep architectures, but unlike standard deep models, it offers bio-inspired mechanisms and relates more to the processes in the brain.

2 Our models

2.1 Robot MNS

First of our models that are building high-level multimodal representations was our robotic mirror neuron system (MNS) model (Rebrová et al., 2013). It consists of several modules: the core and the topmost part is the mirror neuron circuit itself, which is connected to association areas and the low-level modules for execution of movement and gathering of visual information.

In our modeling we assume the sensory-motor links are established between higher level representations, rather than directly between low-level representations of the movement. To encompass the temporal nature of the action sequences (i.e. joint angles of the robot changing in time), the association areas are made of the special type of recurrent self-organizing maps, namely the merge-SOMs (MSOM, Strickert & Hammer, 2005). For our experiments we used the simulated version of the humanoid iCub robot (Tikhanoff et al., 2008) trained to perform three different grasping actions. These are processed by the higher level association area (MSOM) and get self-organized on the resulting maps. We observed that this organization is according to the type of the grasp. Similarly, the higher level association area for vision received information from low-level vision module in terms of joint positions from different viewpoints.
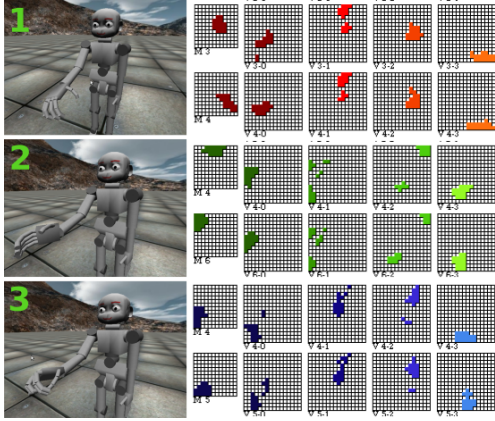
Fig. 2: The left side of the figure shows the examples of iCub's grasping actions of three types and on the right side activations of the trained motor MSOM of 12×12 neurons and from the visual MSOM of 16×16 neurons for each view-point, both binarized with $k = 16$.

This MSOM gets organized primarily according to the viewpoints and secondarily according to the grasp types (Fig. 2).

During the production of the movement, the motor information and the visual information from the self-observation perspective are binarized via the k-WTA method[1] and associated bidirectionally using the BAL (Bidirectional Activation-based Learning) neural network (Farkaš & Rebrová, 2013). We assume that the robot observes another robot producing the same actions and creates visual representations of those actions from viewpoints and associates them with the motor representations as well. After the model is trained, the action observed by the robot elicits the motor representation of the action. These are projected to the topmost module together with visual information, where the motor information helps to form the view-independent representations (Fig. 3).

In this previous work we proposed the BAL model, but soon found out that it was unable to learn one-to-many mappings required to associate different visual representations from different viewpoints with one motor representation. With a motivation to overcome this problem and also to provide a universal account on learning in the brain, the UBAL (Universal Bidirectional Activation-based Learning) model was born.

2.2 UBAL

The fruit of our research endeavors of cognitive modeling is a novel neural network model UBAL (Malinovská et al., 2019). It represents an alternative to classical, effective, yet biologically implausible error-

---
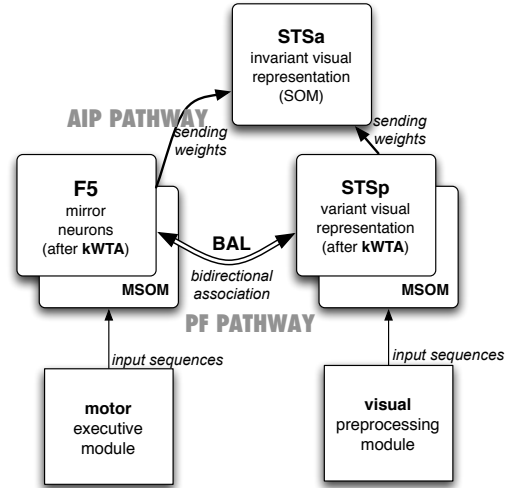
Fig. 3: The robot MNS model (Rebrová et al., 2013).
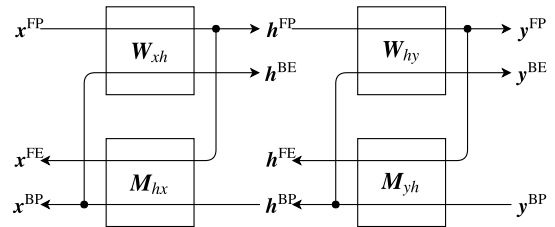
Fig. 4: Schematic deptions of UBAL. The neural activation in the model propagates bidirectionally (F and B) between layers x and y via weights W and M. Activation states computed by the network are also propagated backwards via *echo* connection denotes with E.

backpropagation learning. As well as its predecessor BAL, inpired by GeneRec model (O'Reilly et al., 2012), UBAL is heteroassociative network that does not learn by gradient descent. Unlike any other model UBAL maintains separate weight matrices for two different activation propagation directions between inputs and outputs (Fig. 4). As in the brain the signal only travels through the axons of the neurons in one way via separate synaptic weights. Additionally, UBAL enhances contrastive Hebbian learning rule with an internal echo mechanism enabling self-supervised learning. Therefore, it is able to master various tasks such as association (memory), denoising and classification based on its hyperparameter setup.

Since UBAL approaches any problem as a bidirectional heteroassociation, it has intriguing emergent properties, such as generation of patterns while trained for classification (Malinovská & Farkaš, 2021) as shown in Fig. 5. Such a neural network model can also be understood as a model cognition and the representations it creates on the input level without the stimulation input can be seen as prototypes of the learned categories.
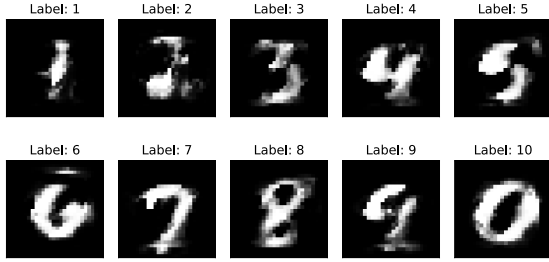
Fig. 5: Projections of learned digits from the MNIST dataset generated by UBAL.
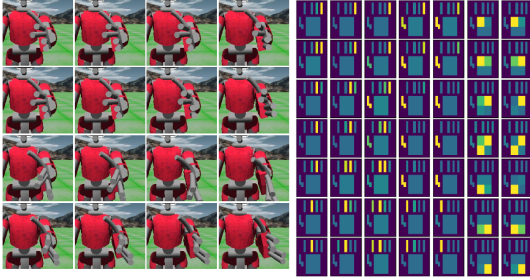
Fig. 6: Proprio (left) and tactile (right) self-organizing map-based association area representations.

2.3 Proprio-tactile associations

In a more recent work within the modular hierarchical modeling paradigm (Malinovská et al., 2022) we explored the phenomenon of the emergence of the body schema which built form our early infancy. Body representations have a multimodal nature that foremost comprises somatosensory, i.e. tactile and proprioceptive information, coupled in a hierarchical fashion in high-level multi-modal representations. In line with the cognitive robotics paradigm we proposed a simple connectionist model build from hierarchically connected neural networks that learns the proprioceptive-tactile representations for the humanoid iCub robot.

This model, along with the previous MNS model utilizes self-organizing maps (SOM, Kohonen, 1997) as association areas (Fig. 6, the k-WTA mechanism and the UBAL model at the topmost level. It has shown a quite good ability to predict touch and its location from proprioceptive information. Due to the generative properties of UBAL mentioned above we could also observe how the model predicts the body-configuration based on the information from the tactile modality, even though the association is many-to-one rather then one-to-one. This intriguing properties along with other aspects such as the quality of the high-level representations emerging on the hidden layer of the model are worth further investigation and modeling and are subject to ongoing research.
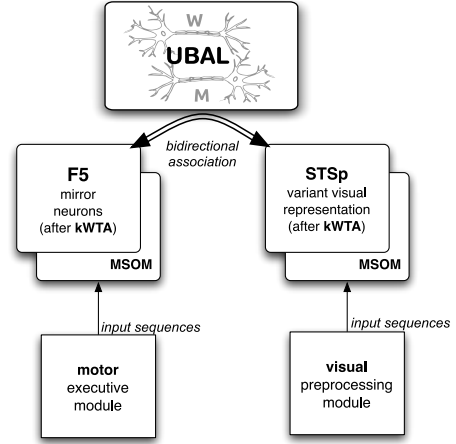
Fig. 7: The robot MNS model (Rebrová et al., 2013).

2.4 Robot MNS: a revival

Our path towards robot mirror neurons or robot common coding model is still unraveling. Our current effort is to revisit the old MNS model and replace the BAL module with the UBAL module which would in turn also replace the topmost association area of the model. Progressing with the current knowledge on the UBAL model we can also assume interesting phenomena if we turn the prediction task the other way round and let our model predict what would be seen when the action is executed.

The modular neural architecture (Fig. 7 is composed of low level execution modules that can be various different control models[2] and vision processing module which could be a deep neural network for extraction of the observed joint positions. The core of our model is representing the movements in motor (proprioceptive) and visual modalities as patterns in a topologically organized association area module utilizing the MSOM model (Strickert & Hammer, 2005). On the top these representations are connected via the UBAL model, capable of connecting multiple different visual representations with one motor representation (robot's own motor repertoire).

A vital task for this research is to collect an adequate sample of data from the robot platform and also to extend from the iCub robot to other platforms. In our case we are working with the small humanoid NICO robot (Kerzel et al., 2017), which we have also in physical form. The model that allows the robot to connect its own movement with the observed one could be beneficial for human-robot interaction and smooth implementation of robot's ability to imitate the human partner. The core idea behind is to use the hierarchical neural model that will allow transfer on the conceptual level, rather then on the concrete effector level, which is also

---

[2]Modules generating joint angle sequences for the robot such as an inverse kinematics module or RL-trained neural network such as the one used in the previous model

a known property of the mirror neurons (Rizzolatti & Sinigaglia, 2010). The aim for the next modeling would be to have the way to retrieve the robot's movement patterns from the association areas. Such model, connected to a helpful software processing the human moving in real time such as the MediaPipe (Lugaresi et al., 2019) can be then used for the interaction of the robot with the human in imitation learning and collaboration tasks.

3 Conclusion

Cognitive robotics aims at studying cognitive capacities of humans via building them in humanoid robots. Capabilities such as action understanding or body schema building are a result of hierarchical multi-modal processing and therefore we claim it is opportune to build them via modular hierarchical neural network architectures with use of biologically plausible neural learning mechanisms such as self-organizing maps and bio-inspired association models (UBAL). In our future work we aim at further developing and testing the model for action understanding and use it in the context of human-robot interaction in imitation and collaboration tasks.

Acknowledgement

References

Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22(04):577–660.

Farkaš, I. & Rebrová, K. (2013). Bidirectional activation-based neural network learning algorithm. V *Proceedings of the International Conference on Artificial Neural Networks (ICANN), Sofia, Bulgaria*.

Hinton, G. E. & McClelland, J. L. (1988). Learning representations by recirculation. V *Neural Information Processing Systems*, str. 358–366. New York: American Institute of Physics.

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05):849–878.

Kerzel, M., Strahl, E., Magg, S., Navarro-Guerrero, N., Heinrich, S. & Wermter, S. (2017). Nico - neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction. V *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, str. 113–120.

Kohonen, T. (1997). *Self-organizing Maps*. Springer.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M. & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines.

Malinovská, K. & Farkaš, I. (2021). Generative properties of Universal Bidirectional Activation-based Learning. V *Artificial Neural Networks and Machine Learning - ICANN 2021*, str. 80–83.

Malinovská, K., Farkaš, I., Harvanová, J. & Hoffmann, M. (2022). A connectionist model of associating proprioceptive and tactile modalities in a humanoid robot. V *IEEE International Conference on Development and Learning (ICDL)*.

Malinovská, K., Malinovský, L., Krsek, P., Kraus, S. & Farkaš, I. (2019). UBAL: a Universal Bidirectional Activation-based Learning Rule for Neural Networks. V *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems*, str. 57–62.

O'Reilly, R. C., Munakata, Y., Frank, M., Hazy, T. et al. (2012). *Computational Cognitive Neuroscience*. PediaPress.

Pfeifer, R. & Scheier, C. (1999). *Understanding Intelligence*. The MIT Press, Cambridge, MA.

Rebrová, K., Pecháč, M. & Farkaš, I. (2013). Towards a robotic model of the mirror neuron system. V *Proceedings of the 3rd joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*. Osaka, Japan.

Rizzolatti, G., Fabbri-Destro, M. & Cattaneo, L. (2009). Mirror neurons and their clinical relevance. *Nature Clinical Practice Neurology*, 5(1):24–34.

Rizzolatti, G. & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4):264–74.

Strickert, M. & Hammer, B. (2005). Merge SOM for temporal data. *Neurocomputing*, 64:39–71.

Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L. & Nori, F. (2008). An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator. V *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, str. 57–61. ACM.

---

[3]https://cogsci.fmph.uniba.sk/sskv/

Explainable Addressee Estimation

Iveta Bečková and Štefan Pócoš

Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava
Mlynská dolina F1, 842 48 Bratislava, Slovak Republic
{iveta.beckova, stefan.pocos}@fmph.uniba.sk

Abstract

Addressee estimation is one of the tasks present in the multiparty human–robot interaction (HRI). In this study, we enhance the model used for estimating the addressee by adding an attention module. By extracting attention weights, we can produce user-understandable explanations about the model predictions.

1 Introduction and motivation

The task of addressee estimation, i.e., understanding who the speaker is talking to, is often overlooked, but every social robot must function well in a multiparty interaction. As with many other computer vision tasks, researchers are trying to solve it using deep learning. However, deep neural networks (NNs) are mostly black-box, producing outputs without any explanation. Especially in HRI, by making the robot's behaviour more explainable, we can also make it more trustworthy, contributing to smoother interaction with humans.

In our work, we enhance an existing addressee estimation model by adding multiple attention modules, utilising attention mechanisms to provide a certain level of inherent explainability. By interpreting the attention weights as importance scores, we can produce multiple different explanations of the model outputs.

2 Previous work

Recently, the task of addressee estimation has been studied by Mazzola et al. (2023). To make a fair comparison, we use the same dataset and data pre-processing. The dataset contains short videos captured from the robot's perspective. Each video shows two people in conversation, and the task is to classify it into one of three classes, based on whether the addressee is on the left, on the right, or the robot itself. The estimation works as follows: first, the speaker's pose is extracted from the video, then their face is cropped (based on the location of the head). The face images and pose vectors are processed separately, then concatenated and fed into a recurrent NN, which produces the final classification of the entire sequence.

3 Adding attention module

There are multiple ways of incorporating attention (Brauwers and Frasincar, 2023). Using a vision transformer instead of a convolutional NN results in pixel-wise attention weights of the input images. Next, by replacing the simple concatenation of processed face and pose vectors with an attention module, we can get importance scores for face vs. pose in each time frame. Finally, considering not only a single frame but an entire sequence, we can feed it into a gated recurrent unit network, utilising its hidden state as a query, which is then used to compute attention scores for individual time frames leveraging general attention. Combining these approaches in a single model, we can get two different importance weightings in real-time and an additional one at the end of each sequence.

4 Crafting explanations

Extracted importance scores serve as a basis for crafting human-understandable explanations. Pixel-wise attention weights are used to overlay with the input image to visualise important parts of the image. Weights of face vs. pose can also be visualised in real-time to highlight which of the data streams is more relevant for the network. The weighting of time frames is used to determine which part of the interaction was the most crucial for estimating the addressee. That information is communicated to the human participants using the robot's speakers.

5 Results

The resulting network with three different attention blocks is on par with the state-of-the-art model for this task. However, by using attention, we are also able to provide real-time feedback to the user about which part of the video is the most salient, and the relative importance of face vs. pose information. Our model can also provide an estimate of which part of the conversation was the final classification based upon — the beginning, the middle, or the end of the conversation.

Acknowledgement

References

Brauwers, G. and Frasincar, F. (2023). A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3279–3298.

Mazzola, C., Romeo, M., Rea, F., Sciutti, A., and Cangelosi, A. (2023). To whom are you talking? a deep learning model to endow social robots with addressee estimation skills. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10.

Learning distributed representations in a model of associative memory

Matej Fandl, Martin Takáč

Centre for Cognitive Science
Faculty of Mathematics, Physics and Informatics
Comenius University in Bratislava
{matej.fandl,martin.takac}@fmph.uniba.sk

Abstract

Associative memories are content-addressable storage devices, the computational models of which are theoretically and practically useful. Modern Hopfield networks are a family of associative memory models with high capacity and fast memory retrieval. This paper presents design principles and early experimental evaluation of a model of associative memory with distributed representations of memories inspired by this family of models.

1 Introduction

Associative memory is a system capable of storing relations between seemingly unrelated items (Camina & Güell, 2017; Suzuki, 2008). It is integral to the everyday functioning of humans and untangling the associative mechanisms of the human brain is a fruitful and active research area (Suzuki, 2008; Wang & Cui, 2018). Artificial neural networks proved to be a promising direction in modeling associative memories both historically (Hopfield, 1982; Kanerva, 1988) and in present (Demircigil et al., 2017; Krotov & Hopfield, 2021, 2016; Millidge et al., 2022; Ramsauer et al., 2021).

Hopfield network (Hopfield, 1982) is a well-known and well-studied (Amit et al., 1987) discrete recurrent neural model of associative memory. It is important in the field but is hindered by low storage capacity, exceeding which can result in incorrect memory retrieval (Hopfield, 1982; Ramsauer et al., 2021).

Dense Associative Memory (Krotov & Hopfield, 2016) – a generalization of the original model – solves its capacity problems. It inspired a new line of research into a family of models known as Modern Hopfield Networks. This led to even higher storage capacity (Demircigil et al., 2017), continuous version of the model (Ramsauer et al., 2021), and arbitrarily deep models (Krotov, 2021).

Training options include storing all exemplars as network weights or using backpropagation of error algorithm (Krotov & Hopfield, 2016). The former method is fast training-wise, but each stored pattern increases the space and time complexity of the model. The latter method is known for being slow and not biologically plausible (Lillicrap et al., 2020).

There are many areas where associative memories with large storage capacity are useful – for an overview see for example Krotov & Hopfield (2021). A model with reduced computational complexity is desirable because it can be trained and deployed in scenarios with limited resources, such as in consumer-grade electronic devices.

In this paper, we report preliminary results from exploring a new model of associative memory. We aim to solve the computational complexity issues by reducing the network size and therefore reducing the computational complexity of pattern retrieval, but utilizing fast, biologically inspired learning. We build on the continuous Modern Hopfield network by Ramsauer et al. (2021). We update the activation dynamics so that it works in tandem with a competitive training procedure that leads to the formation of distributed representations.

2 Our model

Our network consists of one visible layer of d neurons (the network state $\boldsymbol{\xi}$) and one hidden layer of M neurons (as in Krotov & Hopfield (2021)), mainly inspired by the continuous Modern Hopfield networks by Ramsauer et al. (2021). We aim to achieve distribution of labor between hidden units to store a training set $\boldsymbol{X} = \{\boldsymbol{x_i}\}_{i=1}^{N}$, where $\boldsymbol{x_i} \in [-1, 1]^d$ into a matrix of memory vectors $\boldsymbol{W} = \{\boldsymbol{w_i}\}_{i=1}^{M}$ such that $M \ll N$.

2.1 Activation dynamics

The model builds reconstructions from parts shared across the training patterns. These parts – *the memory fragments* – are stored as individual memory vectors $\boldsymbol{W}$. Querying the network means setting ξ to values of a probe and performing one update step with the state update rule

$$\boldsymbol{\xi}^{new} = \mathrm{clamp}_{[-1,1]}\left(\boldsymbol{W}\boldsymbol{s}^{\circ\beta}\right), \qquad (1)$$

$$\boldsymbol{s} = \{s_i\}_{i=1}^{d}, \qquad (2)$$

$$s_i = \mathrm{cossim}\left(\boldsymbol{w_i} \circ \boldsymbol{w_i}, \boldsymbol{w_i} \circ \boldsymbol{\xi}\right), \qquad (3)$$

Algorithm 1 One epoch of training procedure.

> **for** $pattern$ **in** $allPatterns$ **do**
> $residue \leftarrow pattern - reconstruction$
>   **for** $iteration \leftarrow 0$ **to** $maxIterations$ **do**
> $cw \leftarrow closestWeightsTo\,(residue)$
> $updateWeights\,(cw, residue)$
> $residue \leftarrow residue - cw \cdot s^{\circ\beta}$
>   **if** $norm\,(residue) < threshold$ **then**
>   **break**
>   **end if**
>   **end for**
> **end for**

where $\boldsymbol{W}$ is a column matrix of memory vectors, $\boldsymbol{s} = \{s_i\}_{i=1}^{M}$ is a vector of similarities of memory vector $\boldsymbol{w_i} = \boldsymbol{W}_{:,i}$ with the current state $\boldsymbol{\xi}$, β is the inverse temperature and $^\circ$, $\circ$ are Hadamard (element-wise) exponentiation and product operators, respectively. The components of $\boldsymbol{\xi}^{new}$ are clamped to a range between -1 and 1.

The similarity function (3) achieves that zero-valued components in a memory vector are not contributing to the similarity and the effect of close-to-zero components is significantly reduced. This way, the amount of similarity indicates how much $\boldsymbol{\xi}$ is locally similar to the relevant part encoded by a particular memory vector. The value of s_i is a value between -1 and 1, where value 1 means that the $\boldsymbol{w_i}$ and $\boldsymbol{\xi}$ have the same direction, value -1 means that they are opposite and 0 means there is no correlation at all - the vectors are orthogonal.

2.2 Training procedure

We propose a sequential competitive training procedure governed by an iterative residue mechanism described in Algorithm 1.

A residue[1] is maintained throughout all iterations for a single pattern. Its initial value is set to the difference between the current training pattern and its reconstruction (1).

Several iterations [2] follow where a memory vector closest to the actual residue is chosen with

$$\arg\max\left(\mathrm{cossim}\left(\boldsymbol{w_i} \circ \boldsymbol{w_i}, \boldsymbol{w_i} \circ \mathbf{residue}\right)\right). \quad (4)$$

If the closest neurons' similarity does not exceed the *similarity threshold* a second search pass is performed that maximizes a scaled similarity:

$$\arg\max\left((1 - \phi_i)\,\mathrm{cossim}\left(\boldsymbol{w_i} \circ \boldsymbol{w_i}, \boldsymbol{w_i} \circ \mathbf{residue}\right)\right), \quad (5)$$

---

[1] The missing parts in reconstruction.

[2] The upper bound on the number of iterations is M, but a lower amount is desirable to lower the number of updates.

where ϕ_i is the amount of learning that already happened for each neuron. A neuron can be selected only once for a single training pattern in one epoch. The winner is updated to represent the residue better (6).

Weight changes are computed with the update rule

$$\boldsymbol{w_i}^{t+1} = \boldsymbol{w_i}^{t} + \alpha_i\left(\boldsymbol{v} - \boldsymbol{w_i}^{t}\right), \quad (6)$$

$$\alpha_i = \begin{cases} \alpha^I s_i\left(l + (1-l)\left(1-\phi_i\right)\right) & \text{imprint,} \\ \alpha s_i\left(l + (1-l)\left(1-\phi_i\right)\right) & \text{otherwise,} \end{cases} \quad (7)$$

where $\boldsymbol{v}$ is the target vector, α_i is a neuron-specific learning rate, α^I is a value higher than α that is used in case we want to imprint the current pattern and ϕ_i is the degree to which a neuron already learned. We set $\alpha^I \gg \alpha$ and use it in case the neuron is selected in the second pass. The parameter l is the learning signal leak. It ensures that neurons with full learning capacity ϕ_i update a little even when selected for learning. Its value is usually set to a number close to 0.

The amount of learning for each neuron is tracked as $\phi_i \in [0, 1]$. The more a neuron updates, the slower it learns and the less likely it will be selected for further updates. This lets other (less developed) neurons represent the missing parts.

The updated memory vector scaled by its similarity is subtracted from the residue at the end of the iteration and the algorithm continues with the next neuron (or stops in case the stopping criteria are met). This leads to neurons learning the relevant parts and working collaboratively during the reconstruction phase.

3 Preliminary results

We performed an early set of experiments in which we attempted to store a subset (1000 exemplars, 100 from each class) of the MNIST (LeCun et al., 2010) dataset of handwritten digits into our network with 100 hidden units. We trained the model for 100 epochs with training parameters $\alpha = 0.3$, $\beta = 1$, similarity threshold of 0.2 (used in determining whether a second search pass should occur – see (5)), learning signal leak $l = 0.001$ and maximum iterations per training pattern set to 10. After training we evaluated reconstruction quality with $\beta \in \{1, 3, 6, 9\}$. The training led to the average cosine similarity of clean probes from the training set with their reconstructions of 0.946.

The effect of noise with $\mu = 0$ and varying σ and setting a varying portion of components to 0 can be seen in Fig. 1 and Fig. 2, respectively. In the case of noise, $\beta = 3$ leads to the best reconstruction quality. Higher σ leads to slow performance degradation. Incomplete patterns are more challenging – the performance drops more rapidly and reconstructing with $\beta = 1$ outperforms $\beta = 3$ for a zeroed ratio of more than 30 %. Fig. 3 shows the effect on an example training pattern – an exemplar of number 3. Visual examination supports the
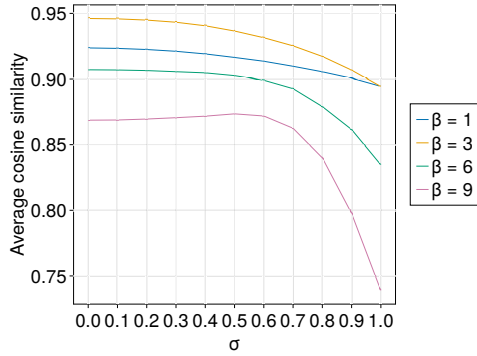
Fig. 1: The effect of noise with $\mu = 0$ and varying σ on reconstruction quality.
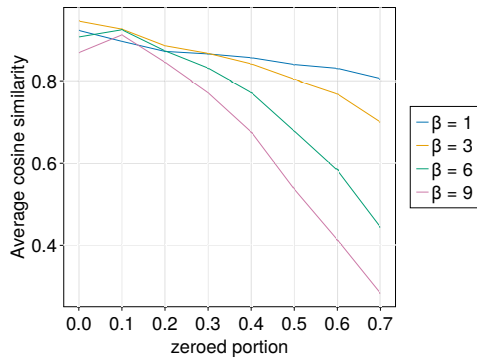
Fig. 2: The value of incomplete patterns on reconstruction quality. Abscissa shows the ratio of components set to 0.

Fig. 3: An example of reconstructions for noisy and incomplete probes of the same pattern – an exemplar of number 3. The noise σ values used (top to bottom) are 0, 0.2, 0.5, and 1. The masked portion ratios were 0.2, 0.5, and 0.7. Color coding: red – negative values, blue – positive values, gray – values close to zero.

results seen in numbers – higher resiliency to noise than incomplete probes.

The resulting weights (Fig. 4) are diverse – two are memory fragments containing the background shared across all training exemplars[3], most are parts of the digits without any background, and some look like complete digits without background. Some neurons did not learn.

4 Conclusion

We introduced a model of associative memory inspired by continuous Modern Hopfield networks aiming to avoid the need to explicitly store all the training patterns as separate memory vectors. The model utilizes a constructive approach to memory retrieval and uses a training procedure designed to support the formation of distributed representations.

Preliminary results show that the model learns distributed representations (lower values of β yield better results) and the reconstructions are identifiable as mem-

---

[3]The positive-valued background memory fragment contributes to reconstruction with component values reversed by its negative similarity s_i.

Fig. 4: Memory vectors from epoch 100. Color coding: red – negative values, blue – positive values, gray – values close to zero.

bers of the target category. The reconstruction quality is more resilient to noise than to incomplete patterns. Visualizing the weights explains why – the resulting memory fragments encode local information and the relation between the known and unknown (zeroed) parts of the probe is easily lost.

Many neurons did not seem to learn anything, which gives the opportunity to either shrink the network size for this dataset or the possibility of future training without catastrophic interference. The latter is a prospective future research direction – along with uncovering the causal effects of hyperparameters on the formed distributed representations, understanding the effect of network size on the reconstruction quality, increasing resiliency to distorted patterns, increasing reconstruction quality, dynamic growth of the network, and experimentally and theoretically evaluating the storage capacity of the network

Acknowledgement

References

Amit, D. J., Gutfreund, H. & Sompolinsky, H. (1987). Statistical mechanics of neural networks near saturation. *Annals of Physics*, 173(1).

Bienenstock, E. L., Cooper, L. N. & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 2(1).

Camina, E. & Güell, F. (2017). The neuroanatomical, neurophysiological and psychological basis of memory: Current models and their origins. *Front Pharmacol*, 8.

Demircigil, M., Heusel, J., Löwe, M., Upgang, S. & Vermet, F. (2017). On a model of associative memory with huge storage capacity. *Journal of Statistical Physics*, 168(2).

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8).

Kanerva, P. (1988). *Sparse Distributed Memory*. MIT Press.

Krotov, D. (2021). Hierarchical associative memory. *ArXiv preprint*, abs/2107.06446.

Krotov, D. & Hopfield, J. J. (2016). Dense associative memory for pattern recognition. Lee, D. D., Sugiyama, M., von Luxburg, U., Guyon, I. & Garnett, R. (zost.), V *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*.

Krotov, D. & Hopfield, J. J. (2021). Large associative memory problem in neurobiology and machine learning. V *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

LeCun, Y., Cortes, C. & Burges, C. (2010). Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2.

Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J. & Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6).

Millidge, B., Salvatori, T., Song, Y., Lukasiewicz, T. & Bogacz, R. (2022). Universal hopfield networks: A general framework for single-shot associative memory models. Chaudhuri, K., Jegelka, S., Song, L., Szepesvári, C., Niu, G. & Sabato, S. (zost.), V *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, vol. 162 z *Proceedings of Machine Learning Research*. PMLR.

O'Reilly, R. C. & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press, 1st. vyd.

Ramsauer, H., Schäfl, B., Lehner, J., Seidl, P., Widrich, M., Gruber, L., Holzleitner, M., Adler, T., Kreil, D. P., Kopp, M. K., Klambauer, G., Brandstetter, J. & Hochreiter, S. (2021). Hopfield networks is all you need. V *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Rizzo, R. & Chella, A. (2006). A comparison between habituation and conscience mechanism in self-organizing maps. *IEEE transactions on neural networks*, 17(3).

Sacouto, L. & Wichert, A. (2023). Competitive learning to generate sparse representations for associative memory. *Neural Networks*, 168.

Suzuki, W. A. (2008). Associative learning signals in the brain. *Prog Brain Res*, 169.

Wang, J.-H. & Cui, S. (2018). Associative memory cells and their working principle in the brain. *F1000Research*, 7.

Low-level causality in a robotic sensorimotor behavior[*]

Miroslav Cibula, Igor Farkaš

Faculty of Mathematics, Physics and Informatics
Comenius University Bratislava
cibula25@uniba.sk, igor.farkas@fmph.uniba.sk

Abstract

Humans and other higher animals are able to observe and learn the causal relationships between the actions they take and their perceptual consequences in the environment. This concept has inspired the field of robotics with an argument that understanding causal relations as a fundamental ability is a prerequisite for building advanced AI systems with common sense. In this work, we study and evaluate low-level causal mechanisms related to a robotic arm that learns the sensorimotor dependencies as well as the effects of its motor actions on the environment.

1 Introduction

Observing and learning causal relations in an environment is essential to human cognition (Gerstenberg and Tenenbaum, 2017). Thanks to this ability, humans can form intuitive theories from multiple observations and use them to predict the environment behaviour in response to their actions (Gerstenberg and Tenenbaum, 2017). This common sense understanding includes the knowledge of intuitive physics, a key ingredient of early cognitive development (Lake et al., 2016).

In this paper, we take inspiration from causal learning in humans (Lombard and Gärdenfors, 2017) and apply it in the field of robotics (Hellström, 2021). We focus on a low-level approach using a robotic arm in a simulated environment, where the arm (an artificial agent) performs random actions and learns by observing their subsequent effects. This process is implemented by training two complementary models that implement learned knowledge. Then, we use methods to interpret that knowledge in terms of causal relations.

2 Problem formulation

We understand a low-level causality as a (transition) function $\mathcal{T}\colon [s(t), a(t)] \mapsto s(t+1)$, where $s(t), s(t+1) \in \mathcal{S}$ are the current (pre-action) and the next (post-action) state of the environment, respectively, from a state space $\mathcal{S}$, and $a(t) \in \mathcal{A}$ is an action from an action space $\mathcal{A}$ performed at time t. We refer to $\Delta s_a(t+1) =$

$s(t+1) - s(t)$ as an effect of action a, reflected in changes of some features of the environment state. Additionally, we understand $\mathcal{T}$ as a low-level intuitive theory encapsulating the accumulated knowledge of causal relations in a given environment.

3 Methods

The methods we use involve two components: causal learning and subsequent knowledge extraction. Both are described below.

3.1 Causal learning

Causal cognition in robots has been proposed to include a range of categories, varying in terms of complexity (Hellström, 2021). Here, we focus on the lower end of this spectrum and discuss low-level causality regarding two categories: sensorimotor self-learning (C1) and learning the consequences of one's own actions on objects in the environment (C2). Both categories naturally imply embodied knowledge, reflected in arm geometry and kinematics.

Learning of both categories of causality required first offline collection of observation data using motor babbling in a simulated environment for which we used myGym toolkit (Vavrečka et al., 2021). In each step, the agent (robotic arm) executes a randomly selected action a and observes a new state $s(t+1)$. Motor babbling is a natural process observed in infants during their first months. In the case of interaction with objects, the concept of intuitive physics becomes relevant.

In the case of C1, the arm performs motor babbling and records its joint configuration and Cartesian effector position before and after the execution of action a. In the case of C2, an object is added on the table in the simulated environment, and the arm has the possibility to interact with it using constrained motor babbling. During an episode, the agent observes potential changes in position, rotation and other defined features of the object, arm and environment in response to the arm actions.

Observations before and after executed action along with the action vector (as per definition of $\mathcal{T}$) collected from the data generation stage are subsequently used for

---

the training of two standard models in robotics:

- *forward model* FM: $[s(t), a(t)] \mapsto s(t+1)$ and
- *inverse model* IM: $[s(t), s(t+1)] \mapsto a(t)$.

We implemented both models using supervised feed-forward neural networks. FM is a well-defined, causal model trying to approximate $\mathcal{T}$. In contrast, IM is non-causal since it reverses effects and causes in time. The FM contributes to causal knowledge and provides the agent with predictions of the next state caused by a current action. A trained FM can be utilized for chained inference of several steps ahead (mental simulation) if provided with a sequence of actions.

3.2 Knowledge extraction

Trained causal models can be analyzed by extracting information about the original environment and a learning session. Our primary focus is on analyzing feature importance, which allows us to highlight state features that cannot be manipulated by the agent actions and thus can be removed, hence reducing the dimensionality of the state space.

Recent related work by Lee et al. (2021), which served as an inspiration, focused on determining relevant state features by conducting intervention on one feature at a time and testing whether the same policy execution led to successful task completion or not. This way, causal dependencies were found.

On the contrary, we do not study causality by direct interaction with an environment but by using trained causal models as proxies containing this information. Using the learned FM we can determine the relevance of state features in relation to action features by analyzing the feature importance within the FM.

To do this, we take a sample of generated causal data and explain the prediction made by the FM for each instance using Deep SHAP method (Lundberg and Lee, 2017). This method uses attribution rules of DeepLIFT (Shrikumar et al., 2017) analysis technique to approximate Shapley value of each input feature in relation to an output feature. Shapley value represents an input feature contribution to the output feature prediction. This way, we can determine a contribution of each action variable to each state variable. Local explanations are then aggregated across the data sample.

The resulting global contribution heatmap generated from the FM trained on the C2 task is shown in Figure 1. The task consisted of a robotic arm randomly switching its magnetic endpoint. Upon turning the magnet on, the arm navigated to the cube on the table, picking it up and randomly manoeuvring with it in the space. Here, the y-axis denotes the action of each joint and a magnetic endpoint of the arm. The x-axis then contains defined environment state features. The colour of each square corresponds to the magnitude of contributions of action features to the state features averaged across 200
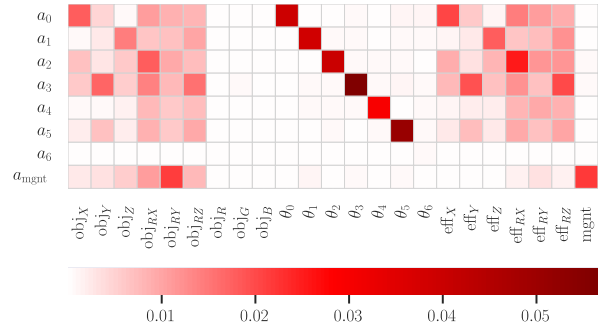
Fig. 1: A heatmap generated by Deep SHAP method on the FM showing the contributions of action features' to the state features (object position and rotation, joint angles, effector position and rotation, and magnet state).

samples. The figure shows, for instance, that joint 6 is not used in the sampled observation data. In addition, the colour of an object ($\mathrm{obj}_{\{R,G,B\}}$) is irrelevant in this case as no action can affect it, and thus could be removed (or ignored). On the other hand, all action features affect most object features. This low-level knowledge can be useful for causal analysis at higher levels.

References

Gerstenberg, T. and Tenenbaum, J. B. (2017). Intuitive theories. In *The Oxford Handbook of Causal Reasoning*. Oxford University Press.

Hellström, T. (2021). The relevance of causation in robotics: A review, categorization, and analysis. *Paladyn, Journal of Behavioral Robotics*.

Lake, B., Ullman, T., Tenenbaum, J., and Gershman, S. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences*.

Lee, T., Zhao, J., Sawhney, A., Girdhar, S., and Kroemer, O. (2021). Causal reasoning in simulation for structure and transfer learning of robot manipulation policies. In *Int. Conf. on Robotics and Automation*.

Lombard, M. and Gärdenfors, P. (2017). Tracking the evolution of causal cognition in humans. *Journal of Anthropological Sciences*.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Int. Conference on Neural Information Processing Systems*.

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features through propagating activation differences. In *Int. Conf. on Mach. Learning*.

Vavrečka, M., Sokovnin, N., Mejdrechová, M., and Šejnová, G. (2021). MyGym: Modular toolkit for visuomotor robotic tasks. In *Int. Conf. on Tools with AI*.

Enlightening Metaphors of Artificial Intelligence

Jiří Wiedermann

Centrum Karla Čapka pro výzkum hodnot ve vědě a technice
Ústav informatiky AV ČR
Pod Vodárenskou věží 2, 182 07 Praha, Česká republika
Email: jiri.wiedermann@cs.cas.cz

Abstract Metaphors offer a great way to understand new concepts. In this essay, we introduce several metaphors by prominent AI luminaries that have recently appeared in the media and are used to enlighten the concept of AI in general and LLMs in particular. None of these metaphors completely captures the essence of the notion at hand, but each illuminates some aspects. It is hoped that the union of all intuitions behind all these metaphors will result in an original mental image leading to a new apprehension and appreciation of the respective technologies and their potential. In the accompanying talk, we will discuss the weak and strong points of the individual metaphors.

What is it: it's wise, but not alive?[1]
(J. Wiedermann & J. van Leeuwen, 2023)

The Metaphors

With the advent of large language models, metaphors have emerged that attempt to explain in an accessible way how these models, or artificial intelligence as such, work and what their purpose is. Like all metaphors, they limp along, but they do succeed in enlightening, often in an entertaining way, at least some of the salient features of artificial intelligence. For the sake of interest, we will list the most interesting ones without pointing out their shortcomings. It is up to the kind reader to judge their suitability and appreciate their originality. In the following overview, with each metaphor, in parentheses its author is listed, along with its brief description.

The metaphor of a stochastic parrot (Emily M. Bender, American linguist). In this context, "stochastic" means a process that has a random, unpredictable evolution, and is applied to large language models. A stochastic parrot glues together parts of text observed in large training data according to statistical regularities of their possible combinations but without any relation to the meaning of the text.

The metaphor of a blurred computer image of the entire Internet (Ted Chiang, American science fiction

writer). According to this metaphor, a large language model can be seen as a blurred image of the whole web. Prompt determines the approximate part of the image in which to look for the answer, and the details of the answer are "computed" from the blurred part of the image using the statistical laws mentioned in the case of the stochastic parrot.

The talking dog metaphor (Terrence Sejnowski, pioneer of computational neuroscience). This metaphor of large language models is given by the following parable:

The following story is about a chance encounter in rural America when a curious driver came across a sign that read, "TALKING DOG FOR SALE." The owner took him into his yard and left him with an old border collie. The dog looked up and said:

"Woof! Woof! Hi, I'm Carl, nice to meet you." The driver was impressed.

"Where did you learn to talk?"

"Language school," Carl said, "I was in a top-secret CIA language program. They taught me three languages: How can I help you? Wie geht es ihnen? Jak se máte?"

"That's incredible," said the driver, "What was your job at the CIA like?"

"I was a field agent and the CIA flew me all over the world. I sat in a corner and eavesdropped on conversations between foreign agents and diplomats who never suspected I understood what they were saying, and I reported to the CIA what I overheard."

"You were a spy for the CIA?" The driver said, increasingly astonished.

"When I retired, I received the Distinguished Intelligence Service Cross, the highest decoration awarded by the CIA, and honorary citizenship for extraordinary service rendered to my country."

The driver was a bit shaken by the encounter and asked the owner how much he wanted for the dog.

"You can have the dog for 10 dollars."

"I can't believe you're asking so little for such a wonderful dog."

"Did you really believe all that CIA bullshit? Carl never left the farm."

Consulting firm metaphor (Ted Chiang). Within this paradigm, we view AI as a technology promoted by its parent companies to increase their profits.

King Midas metaphor (paper clip metaphor) (Nick Boström, Swedish philosopher). Artificial intelligence

---

[1] Generative AI

is an entity that literally, but in the least acceptable way, carries out the wishes of its creator.

Thing on the cloud metaphor (Sam Altman, CEO of OpenAI, which developed ChatGPT). General AI is "that thing that resides on the (computational) cloud, with humans serving as its robotic arms".

The metaphor of an autonomous corporation (Ilya Sutskever, Research Director of OpenAI). The artificial intelligence of the future will be an automated autonomous research, development, and manufacturing organization.

The metaphor of extraterrestrial intelligence (Yuval Harari, Israeli historian and writer). The acronym AI (artificial intelligence) is an abbreviation for "alien intelligence".

The metaphor of a stupid agent (Luciano Floridi, a prominent Italian-British philosopher). Artificial intelligence is the ability to act without intelligence.

Parental metaphor (J. Wiedermann & J. van Leeuwen). The intergenerational relationship of artificial intelligence to future ones, and specifically to humanity, should be like the relationship between parents and their children: when children are young, they are cared for and raised by their parents; when children are grown and parents are old, they are cared for and honored by their offspring.

Sherlock Holmes metaphor (J. Wiedermann). In language, an LLM is akin to a brilliant detective, sifting through a vast sea of information, eliminating improbable explanations, and revealing the hidden truth.

As can be seen, none of the above metaphors is entirely apt to what it is meant to define. The reason, according to Boaz Barak, an Israeli-American professor of computer science, is that advanced artificial intelligence systems are never-before-seen objects with properties not previously possessed by physical objects, as the opening riddle suggests. They are fundamentally new objects that we cannot compare to anything else.

Theory of Mind: Varying Substrates of Social Cognition

Alexandra Dyalee

Faculty of Mathematics, Physics and Informatics
Mlynská dolina 842 48 Bratislava
alexandra.dyalee@mensa.sk

Abstract

In the paper, we firstly provide insight into implicit social brain from multidisciplinary perspective, containing views of psychology, neural substrates and methods of clinical measurement. On the grounds, we outline philosophical consequences. We assess that implicit social brain realization can be as conscious, as unconscious.

Secondly, we focus on a possible altered[1] execution of particular tasks of theory of mind (ToM) in people with autism spectrum disorders (ASD) and Asperger syndrome in particular[2]. Referring to this, we aim to exalt the spectrum of varying possibilities of the cognitive representation of ToM with respect to neurotype.

Finally, we suggest two novel analysis components for effective evaluation of ToM testing tasks, possibly helping us distinguish between particular points within the spectrum of its potential cognitive representations.

1 Definition

Before we expostulate on theory of mind, we stress out the fact that such concept is in literal *a theory*, because we are inquiring into something, that is not directly observable, that is mental states of oneself. By these, we understand for instance intention, knowledge, liking, guessing, pretending, etc. The system inquiring for them as such is important as it can be used to make prediction of behavior of others.

By definition, an individual is said to have a theory of mind if he imputes mental states to himself and others (Premack et al., 1978).

2 Why is it important?

Consider following proposition:

"By means of the theory of mind, one is able to be making instant and continuous assumptions of other people's attitude during interaction,"

and exempt the core principle out of it formally.

---

[1] with respect to neurotypical expectation

[2] we treat Asperger syndrome as a discrete diagnostic category according to DSM-IV

Suppose we have a group $\mathcal{G}$ of people such that for every person is true that either he or she is a part of the group $\mathcal{G}$, or he or she is not a part of the group $\mathcal{G}$. Name singular components (persons) of this group $\mathcal{P}_i$. Denote $st(\mathcal{P}_i)$ an attribute of mental state of a person $\mathcal{P}_i$. Notice people are dependent upon mental states of each other. Therefore, $\forall i \in \{1, ..., k\}$:

$$st(\mathcal{P}_i) = f(st(\mathcal{P}_1), ..., st(\mathcal{P}_k)). \tag{1}$$

Without attempting to prove, we pronounce following statement.

Statement. Regarding the **group** $G(\mathcal{P}_1, \ldots, \mathcal{P}_k)$ of people instead of merely the bunch of individuals $\mathcal{P}_1, \ldots, \mathcal{P}_k$ of the group, to be working effectively both from the internal perspective (group sustainability) and external perspective (inter-group interactions), it is a necessary condition for the singular people $\mathcal{P}_i$, to have an information of the state of each other, denote $inf(\mathcal{P}_i)$, and if this is not met by default, then to be able to communicate it somehow fluently.

Examples of such principle might be seen in collective adaptation, self-organized construction of social hierarchy or the very origination of intuitive universal languages, as are for example non-verbal communication or art. Charm of this might be seen in the fact that individuals of a community can be represented as nodes of an interconnected system (interconnected by assigning mental states to all of the other individuals), that is subject to changes as a whole network.

As we see that a significant fraction of social phenomena is unconscious (Bacha, 2017), then we can infer that to make above mentioned principles based on **default communication of mental states** between individuals of a community performable, theory of mind is necessary, as it is a tool for that spreading of unconsciously generated information about mental states between its individuals.

3 A (conscious) spick of autism

Ability to spread information unconsciously does not have to be fully present in every individual. It is widely discussed that people with autism spectrum disorders, including Asperger syndrome, show some impairments in ToM (Duverger et al., 2007). Despite of this, there is

empirical evidence that people with Asperger syndrome on the highly functional[3] end still can perform well in ToM tasks, and the explanation for this lies in the *conscious* compensatory learning (Senju et al., 2009).

4 Neural substrates

Assume that there is no impairment in ToM functioning in neurotypical people. According to PET and fMRI neuroimaging studies reviewed by Leblanc (2020) we see that ToM is a very complex cognitive function, if even we could be talking about it as about a *single* function.

The core lies in that it is not localized to a single critical region, rather it involves activity in multiple brain regions. Particularly, there was established a suggestion (Leblanc et al,. 2020) that the primate social brain is comprised of the "core" regions of orbitofrontal cortex (OFC), superior temporal sulcus (STS) and temporoparietal junction (TPJ) (Fig. 1, 2).
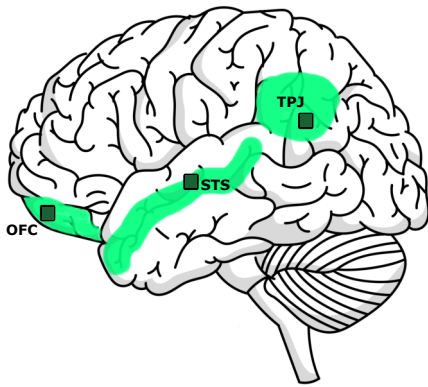
Fig. 1: "core" regions of social brain (lateral view)

Though that some brain regions can be labeled as the most involved, there is still remaining a plurality of constituents undergoing ToM. Also is important to mention explicitly that different ToM tasks activate slightly different brain regions. Moreover, there is not necessarily activated amygdala in every ToM task (Leblanc et al., 2020), out of which we conclude that the role of ToM in navigating social behavior is rather cognitive than affective.

Broadly we make a conclusion that ToM is represented by a specific pattern of communication within a network of interconnected regions, and it is exactly the character of this pattern just as well as the particular network what we are assigning informativeness in regard of observing neural substrates of ToM.

---

[3]with regard to clinical perspective
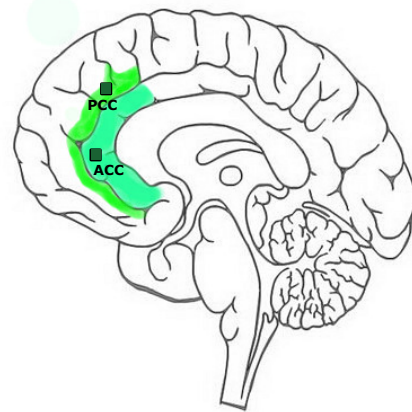
Fig. 2: "core" regions of social brain (sagittal view)

5 Neural substrates in adults with ASD

As mentioned in section (3), ToM mechanism can be also altered, which is particularly a case of people with Asperger syndrome[4]. A recent fMRI study of social animation task performed on 24 people with clinical diagnosis of ASD vs. 24 controls (Chen et al., 2023) suggested that both groups were involved in ToM mechanisms and recruited some shared brain regions during ToM processing (Fig.3),

however there were found additional activations of the middle frontal gyrus (MFG), right superior temporal gyrus (STG) and precuneus in people with ASD (Fig.4).

On the contrary, in control group, there were found additional activations in the left precentral gyrus and right insula (Fig.5).

Both precentral gyrus and right insula have been seen to be the key regions in the *Social Neural Network* (SNN) as well as the human *Mirror Neuron System* (MNS). The precentral gyrus involves the initial step for inferring intentions underlying actions (Gesi et al., 2021). Additionally, the activation of the insula has usually been considered part of the SNN and as a key role in decision making, empathy, and a functional role of social awareness (Hof et al., 2013).

Moreover, as we find concluded by Chen (2023), greater activation of the network of brain regions in ASD group may be attributed to **more effortful processing for adequate task performance**. Particularly on the case of mentalizing, the finding of a positive correlation between greater activation in the right STG and a higher score of the Autism Diagnostic Interview-Revised (ADI-R) in individuals' clinical diagnoses strongly suggests that the poorer the social communication performance in adults with ASD, the more activation in the STG part of their brain. The cor-

---

[4]in cited study, Chen et al. recruited ASD in general with regard to Diagnostic and statistical manual of mental disorders V (DSM-V®)
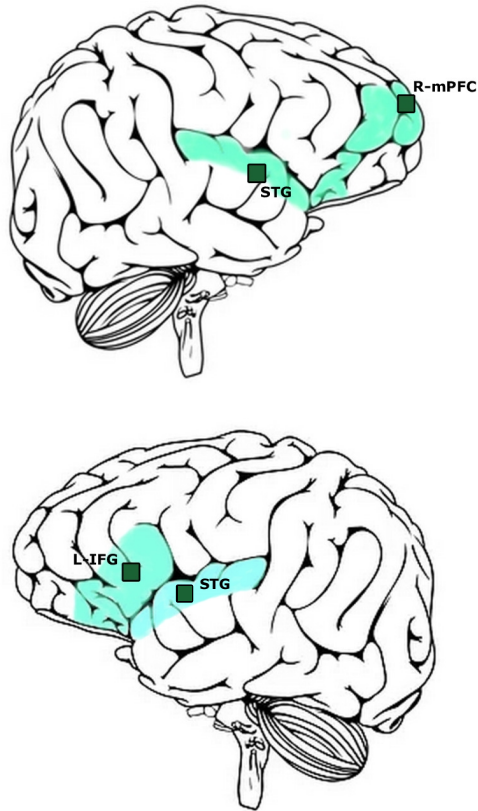
Fig. 3: Social animation task: conjunction of activations

Fig. 4: Social animation task: additional activations in ASD group

relation consistently supports the notion that adults with ASD might spend **excessive effort** on the intentional detecting processing during mentalizing.

6 ToM tasks and measures

Impending to clinical experience, we stress on proper measurement of ToM with an aim to help distinguish between the possible alternative neural substrates of the function. Leblanc et al. (2020) provided a systematic review, screening peer-reviewed literature and relevant test publishers' catalogs, in order to generate an inventory of existing ToM measures that have been used with children under 6 years of age. Note that age is not a limitation in order to measure ToM abilities, yet in younger children less complex testing methods might suffice as we expect an increase in complexity of ToM with age. Importance of development of complex ToM testing methods can be for example seen directly in the need for relevant testing methods for late diagnosis of Asperger syndrome, which is especially frequent in females (Bagriela et al., 2016; Gesi et al., 2021). On the grounds, we suggest following novel analysis components for ToM testing evaluation, potentially helpful specially within the issue of Asperger syndrome diag-

nostics in females.

1. **Measurement of cognitive fatigue.**

   This is an *objective* approach. According to above conclusions, we assume that on the continuum between thought extremes of fully present ToM and fully compensated ToM, cognitive fatigue will be increasing with growing compensation of ToM, indicating present neurodiversity.

2. **Approach analysis.**

   This is a *subjective* approach. Suppose a participant undergoes a particular ToM testing task, as encapsulated by Leblanc (2020). According to the functional distinction between ToM task execution in neurotypical people, that can be concluded to be predominantly unconscious, and neurodiverse people, where the task execution can be altered into a form of conscious[5] analysis aiming for the making of optimal decision, we suppose there will be present qualitatively different approaches toward the task execution in respective groups of participants.

_____
[5]or possibly automatized by learning
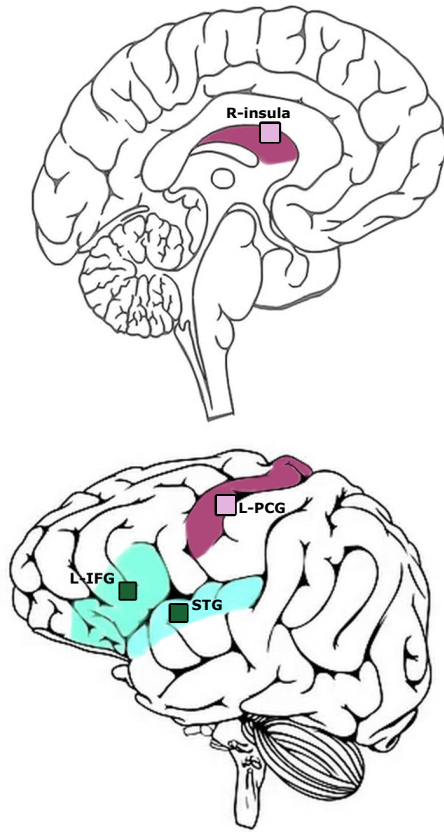
Fig. 5: Social animation task: additional activations in TD group

A method used for above explained analysis is an observed dialogue focused on

(a) presence of participant's need to explain the choice,

(b) presence of an automatic assumption that the interviewer knows the correct answer,

(c) length of answer report.

For successful performing of such analysis, we suppose it is further necessary to devise a manual for evaluation of above-mentioned dialogue, as well as to test it on empirical data. Within this study, we are reducing our attention to the correct frame of theoretical basis.

7 Conclusion

We are pointing to the difference between well-known unconscious realization of ToM tasks in comparison with an option of this functional process to be also conscious, which we can observe in individuals with Asperger syndrome.

On the grounds we conclude there is no single way of how ToM tasks can be executed, which we illustrate

in the Neural Substrates sections ((4) and (5)).

We refer that according to neuroimaging studies (sections (4) and (5)), there is enhanced cognitive load in people with autism spectrum disorders affiliated with ToM tasks execution, particularly with mentalizing, that leads to faster and more intense development of cognitive fatigue in these individuals.

We evaluate that the feasibility of execution of a ToM task by means of multiple neural pathways (sections (4) and (5)) might provide partial answer on the nature of apparent behavioral invisibility of Asperger syndrome in people on its highly-functional end resulting in numerous misdiagnoses. We propose to broaden our focus from examining only the functional presence of particular ToM task execution in observed individuals to the probability of a particular pathway being involved.

Picking up on such, we suggest a novel analysis approach for evaluation of the measure of 'typicalness' of ToM task execution in tested individuals (section (6)). Desire of such approach is to help decrease the error tendency in diagnoses of Asperger syndrome in individuals whose execution of ToM tasks is present, but the process of its realization is altered.

Note that our suggestion relates merely to the theoretical basis of the problem, without any direct aim for a particular design to be drawn. Possible continuation of this study we propose would comprise of assigning and testing particular models.

Simultaneously, our study introduces a conceptual view on what theory of mind is (sections (1) and (2)) and conjectures a starting point on how we could access the unconscious communication between members of one society by the means of a mathematical model (section (2)).

Acknowledgement

References

Apps, M. a Rushworth, M. (2016). The anterior cingulate gyrus and social cognition: Tracking the motivation of others. *Neuron*, 90(4):692–707.

Bacha, C. (2017). Neuroscience and the social unconscious. *Group Analysis*, 50:502–518.

Banker, L.and Tadi, P. (2020). Neuroanatomy, precentral gyrus. *StatPearls [Internet]*.

Bargiela, S.and Steward, R. (2016). The experiences of late-diagnosed women with autism spectrum condi-

tions: An investigation of the female autism phenotype. *J Autism Dev Disord*, 46(10):3281–3294.

Beaudoin, C.and Leblanc, (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, 10.

Carrington, S. a Bailey, A. (2009). Are there theory of mind regions in the brain? a review of the neuroimaging literature. *Hum Brain Mapp*, 30(8):2313–2335.

Chen, T.and Gau, S. (2023). Neural substrates of theory of mind in adults with autism spectrum disorder: An fmri study of the social animation task. *Journal of the Formosan Medical Association*, 122(7):621–628.

Duverger, H. a Fonseca, D. (2007). Theory of mind in asperger syndrome. *Encephale*, 4(1):592–597.

Gesi, C. a Migliarese, G. (2021). Differences in misdiagnosis and delayed diagnosis among adults with autism spectrum disorder with no language or intellectual disability. *Brain Sci*, 11(7):912.

Gu, X. a Hof, P. (2013). Anterior insular cortex and emotional awareness. *J Comp Neurol*, 521(15):3371–3388.

Premack, D. a Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4):515–526.

Senju, A. a Southgate, V. (2009). Mindblind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science*, 325(5942):883–885.

Mechanisms of Contextual Plasticity in Human Sound Localization

Gabriela Andrejková, Stanislava Linková, Norbert Kopčo

Perception and Cognition Laboratory, Faculty of Science, P. J. Šafárik University in Košice
Jesenná 5, 04001 Košice
gabriela.andrejkova@upjs.sk, stanislava.linkova@outlook.com, norbert.kopco@upjs.sk

Abstract

Contextual plasticity is a form of plasticity in sound localization induced by preceding stimulations. It is observed as shifts in responses to a target click stimulus when, on interleaved trials, the target is preceded by an identical distractor coming from a fixed location. Here we present the results of two experiments, one performed in real sound (using speakers) and one in virtual sound (using headphones) environment, evaluated in the context of two models of the neural mechanisms underlying spatial hearing in humans. The first model encodes spatial location by activity of a large population of neurons aimed at accurately encoding the stimulus location. The second model assumes that spatial location is encoded in activity of 4 opponent-processing channels optimized for sound source separation, not localization. The modeling found that performance in the real environment is more aligned with the first model, while performance in the virtual environment is more aligned with the second model, suggesting that listeners use different strategies and/or neural mechanisms when localizing sounds in real vs. virtual environments.

1 Introduction

Contextual plasticity (CP) is a slow adaptive process in spatial hearing induced by a context containing repeated identical stimuli (Andrejková et al., 2023; Kopčo et al., 2007). Little is known about the neural mechanisms underlying the effect (Jeffress, 1948). Here we perform analysis of results from recent series of CP experiments (Piková, 2018; Linková et al., 2022) to examine what mechanisms likely underly the contextual adaptation.

2 Experiments

The experimental methods are briefly summarized here. Two similar experiments were performed in a quiet darkened midsize reverberant room, one in real sound (RE) using speakers and one in virtual sound (VE) anechoic and reverberant environment using headphones. Each trial consisted either of a presentation of a target or an adaptor (50% probability). The target was a 2-ms noise burst (click). The adaptor was a click train consisting of 12 such clicks. Six target locations were used (Fig. 1). Adaptor locations were fixed within a block at 0°, ±45°, or ±90° in Experiment 1, and 0° or ±50° in Experiment 2. Experiments were divided into blocks which kept the adaptor location fixed. Baseline blocks contained no adapter. Eight subjects participated in Exp. 1, nine in Exp. 2. The subjects indicated the perceived target location while seated with their heads supported by a headrest. Response standard deviations were computed for each target in the block to evaluate the models in this study.
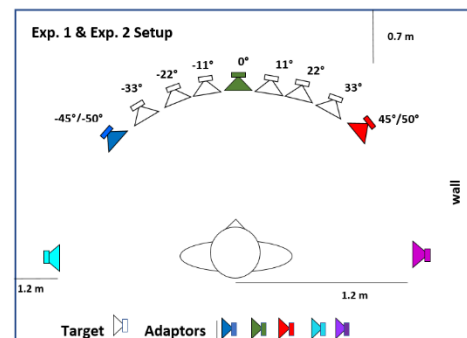
Fig. 1. Setup of experiments.

3 Neural mechanisms

The neural processing can be outlined as follows: Sound waves reaching the ears are first decomposed into frequency channels. Then the signals in each frequency channel from ears are compared in the brainstem and the interaural time differences and level differences are extracted. These two types of cues for azimuthal sound localization are processed separately by specialized neural mechanisms. Two mechanisms for encoding auditory space are considered here:

1. Carlile's model - (Carlile et al., 2001) proposed a model in which a population of units, each tuned to a different spatial location, encodes auditory space. This model assumes that the primary goal of auditory spatial perception is to accurately encode the sound source location, and that the result of adaptation to a repeated presentation of a stimulus from the same location is a fatiguing, causing a suppressed response from the corresponding channel.

2. Lingner's model - (Lingner et al., 2018) introduced an alternative coding model, called a hemispheric balanced model, because it uses independently calculated results for sound localization from both hemispheres, and then combines these results. The model assumes that the goal of the adaptation is to increase separability sources in the region from which most stimuli are presented, resulting in increased discriminability between targets presented near the adaptor.

In the current study it is assumed that in localization this increased discriminability should be accompanied by decreased variance in the localization responses.

4 Results

Fig. 2, upper panels, shows across-listener mean standard deviation (SD) in responses as a function of the target location, separately for the 3 environments (two experiments), and separately for the different adaptor blocks. SD is computed separately for each combination of session, run, target and listener, and then averaged across sessions, runs and listeners. The standard deviations in RE are in the range of 2-4°, with the largest values in positions close to the adaptor for the 45° adaptor and the smallest values in positions far from the adaptor for the -45° adaptor. In VE, errors are larger 3.5-7° and have greater variability.
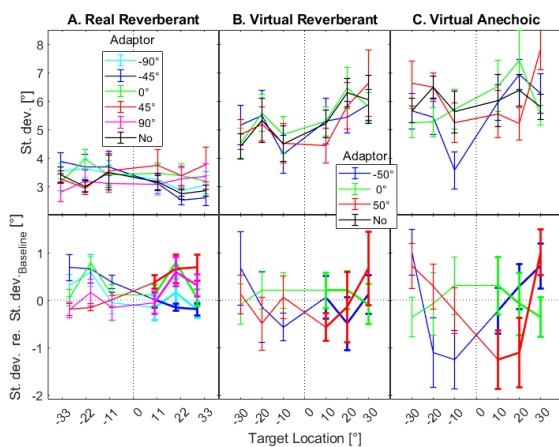
Fig. 2. Standard deviations of raw data (upper row) and a comparison to No adaptor baseline run (lower row) for Exp. 1 (panel A) and Exp. 2 (panels B, C).

The lower panels show the effect of different adaptors by subtracting the no-adaptor data from the respective graphs in the upper panels, and by assuming left-right symmetry in the adaptor effect (thus, e.g., the red and blue graphs are mirrored version of each other).

The adaptors always caused an increase in the response variance in RE (Fig. 2A). ANOVA performed on the Exp. 1 data with factors of *target* (11 to 33°) and *adaptor* (+45, +90°) found a significant main effect of *adaptor* $[F(1, 7)=16.13, p = 0.005]$. ANOVA performed on the Exp. 2 data with the factors of *target, adaptor*, and *environment* (anechoic, reverberant) found a significant

interaction of target and adaptor for the lateral adaptors $[F(2, 16) = 3.75, p = 0.05]$ (the red line for targets at 10-30° in Fig. 2C). This interaction indicates that variance in responses is lowered in presence of the adaptors, however not for the nearest target but for the slightly more distant targets (10 and 20°). Note that no significant interaction with the factor of *environment* was observed. However, visual comparison of panels B and C shows that the effect was observed mostly in the anechoic VE.

5 Conclusion

The increased variance in RE is consistent with the Carlile's model. On the other hand, the decreased variances in the anechoic VE are more consistent with the Lingner's model. Although the standard deviation is high for the nearest target, it quickly decreases afterwards, becoming lower than in the baseline. On average, the presence of the adaptor helped reduce the variability of responses for nearby targets in VE. Thus, it is likely that in RE the listeners use absolute localization, allowing them to map the acoustic cues to an actual sound source location. On the other hand, in VE, listeners might be changing their strategy and using relative localization, e.g., localizing the targets relative to the known location of the adaptor. This interpretation is also consistent with the Carlile and Lingner studies, as the former was performed in RE while the latter was performed in an anechoic VE.

Literature

Andrejková, G., Best, V., & Kopčo, N. (2023). Time scales of adaptation to context in horizontal sound localization. *J. Acoust. Soc. Am.*, *154*(4), 2191–2202.

Carlile, S., Hyams, S., & Delaney, S. (2001). Systematic distortions of auditory space perception following prolonged exposure to broadband noise. *J. Acoust. Soc. Am.*, *110*(1), 416–424.

Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative Physiology and Psychology*, *41*, 35–39.

Kopčo, N., Best, V., & Shinn-Cunningham, B. G. (2007). Sound localization with a preceding distractor. *J. Acoust. Soc. Am.*, *121*, 420–432.

Lingner, A., Pecka, M., Leibold, C., & Grothe, B., A. (2018). A novel concept for dynamic adjustment of auditory space. *Sci. Rep.*, *8*(1), 1–12.

Linková, S., Andrejková, G., & Kopčo, N. (2022). Contextual plasticity in sound localization vs. source separation in real and virtual environments. *Kognice a Umelý Život XX*, 162–163.

Piková, V. (2018). Mechanizmy kontextuálnej plasticity v lokalizacii zvukov. *Bachalor Thesis*, P. J. Šafárik University in Košice.

Semantic Similarity of Action Verbs

Michal Vavrecka

CIIRC, CTU in Prague
Prague 6, Czech Republic
Email: michal.vavrecka@cvut.cz

1 Introduction

We introduce preliminary findings from an ongoing study on the grounding of action verbs. Our approach involves using neural module networks to control a simulated robot, capable of learning multi-step tasks over extended periods by alternating between distinct neural modules. Each module in our multi-policy architecture was trained for a specific subgoal, such as picking, moving, placing, rotating, or wiping, allowing this setup to surpass traditional single-policy algorithms across various manipulation tasks. Following the training phase, we applied a dimensionality reduction technique (tSNE) to visualize the parameters of each module, uncovering correlations between module characteristics and the semantics of the associated actions. These initial outcomes suggest that it is possible to develop action embeddings derived from the semantics distilled from the trained module parameters. Such grounded action verbs could be useful in motion planning, where a planner might substitute similar skills when the desired one is unavailable, or in fostering intrinsic motivation by enabling the generation of new long-horizon tasks based on the semantics of action verbs.

In prior research (1), we introduced two modular multi-policy algorithms, MultiPPO2 and MultiACKTR, designed to enhance performance in various long-horizon tasks by employing distinct policies for each skill with dedicated subgoals. Both MultiPPO2 and MultiACKTR matched the performance of single-policy algorithms in single-step and multi-step translational tasks, but surpassed them in tasks involving object translation, rotation, and path following. In the paper we present, we explore the structure of these modules to test the hypothesis that the similarity of action verbs can be quantified based on the parameters of trained networks. We propose that these network parameters mirror the executed actions and that a low-dimensional projection of these parameters will highlight similarities between actions in a form that is interpretable to humans.

Materials and Methods

Our experiments were conducted using a toolkit designed for training and testing reinforcement learning algorithms, referred to as myGym (2). For the simulati-

ons, we utilized a model of the industrial robot Kuka IIWA. We developed two multi-policy algorithms that are capable of training multiple policies within a single episode and switching between them as the task progresses. Additionally, we created a new set of tasks within the myGym toolkit to evaluate our MultiPPO2 and MultiACKTR algorithms against traditional single-policy algorithms. These multi-policy algorithms consist of separate modules, each representing a specific skill. We applied the tSNE algorithm to their parameters and visualized these in a shared 2D space. Moreover, we analyzed single-step tasks, such as reaching, pressing, switching, and turning, using single-policy algorithms to validate our hypothesis.
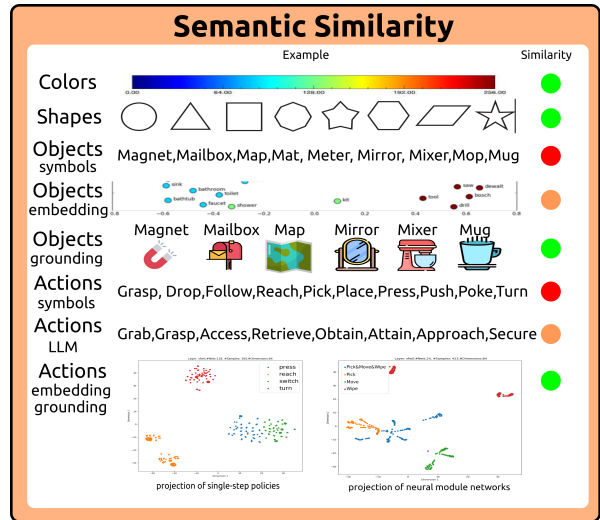
Fig. 1: Semantic similarity comparison table. Object features have semantics, as there is a similarity preserved. The objects preserve similarity only, when they are grounded. The similar situation stands for actions, where symbolic representation does not preserve similariry (red color), the embeddings offers language based similarity (orange color), while only embedded and grounded actions offers full semantics.

Results

The findings confirmed that module parameters vary across subtasks. A correlation exists between the inter-

nal structure of the learned parameters and their functionality within the tasks. If the parameters did not vary among different modules, they would occupy the same subspace in the visualization. The presence of distinct clusters for each module indicates that projecting parameters into the 2D space effectively serves as an embedding of the action verbs. Identical results were observed in both the single-step tasks and the single-policy algorithms.gle-policy algorithms.

Conclusion

We have validated our hypothesis regarding the grounding of action verbs. Similar to how the symbol grounding problem was resolved for static scenes, it is beneficial to represent action verbs within a shared space. The proximity of action verbs in this projection reflects their semantic similarity, as actions of similar types are located close to each other. Creating these embeddings serves several purposes. In motion planning tasks, a predefined set of actions is necessary. If a desired action is not listed, the planner typically fails. However, with grounded actions, the planner can select the action most similar to the desired one and use it as a substitute. This similarity metric is also valuable in the field of intrinsic motivation. Traditional goal-sampling methods are based on distances from known goals, but using action similarity allows for a better interpolation of the manifold of potential skills. This can act as a generator for automatic curriculum development while learning tasks. We plan to further explore this line of research in the future.

Literatúra

[1] M. Vavrecka, J. Kriz N. Sokovnik, G. Sejnova: Modular Reinforcement Learning In Robotic Manipulation Tasks. ICANN2024 **(submitted)**

[2] M. Vavrecka and N. Sokovnin and M. Mejdrechova and G. Sejnova : MyGym: Modular Toolkit for Visuomotor Robotic Tasks. 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), **p. 279-283 (2021)**
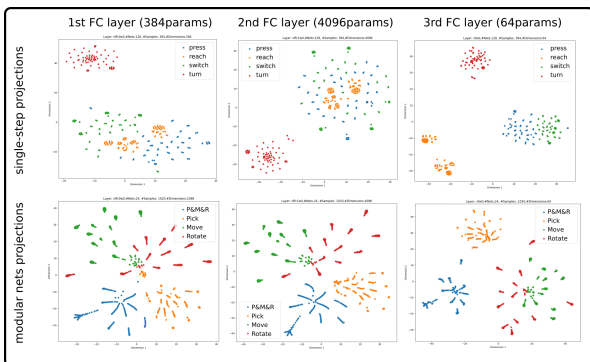
Fig. 2: tSNE projection of trained parameters from value network (ACKTR). The colors represent specific skills trained within subgoals. Each dot represents parameters projection during training (small clusters stands for evolution during training). Each subtask was trained at least 12 times to detect variability.

Approaches to generating arm movements in humanoid robot NICO[*]

Igor Farkaš, Andrej Lúčny, Michal Vavrečka

Faculty of Mathematics, Physics and Informatics, Comenius University Bratislava
Mlynská dolina F1, 84248 Bratislava
{lucny,farkas}@fmph.uniba.sk, michal.vavrecka@cvut.cz

Abstract

The generation of arm movements is a fundamental executive capability associated with robots interacting with an environment. Despite easiness of motor behavior in humans, generating a desired movement in humanoid robots is difficult depending on a number of factors, including the hardware and the task specification. Here we briefly describe various approaches to movement generation with their pros and cons. Generation of smooth and legible movement is often a requirement, especially in case of humanoid robots interacting with humans, which in the future are expected to play an important role in modern society.

1 Introduction

Motor movement is the basic functionality of robots, enabling them to act in the environment. There are two basic categories of motor movement: (1) navigation of a mobile robot in space and (2) object manipulation using robotic arms, affecting the state of the environment.

In our research, we are working with the semi-humanoid robot NICO (Kerzel et al., 2017), which is required to perform movements with the right arm according to the assigned tasks in the context of human–robot interaction (HRI, see Fig.1). The goal of the task is to ensure that the robot performs *legible* movements, i.e. those that better reveal robot *intention*. The legibility of motion is a concept developed in the literature (Stulp et al., 2015). It can be considered as one of the prerequisites for *trustworthy* HRI, which is a relatively new field of research (Kok and Soh, 2020). In this context, humanoid robots represent the most suitable option for successful HRI, mostly because of human tendencies to antropomorphise the robots (Vernon and Sandini, 2024).

2 Approaches to generating arm movements

There exist various approaches toward generation of robotic arm movement, ranging from very rigid engineering ones to the most flexible, machine learning

Fig. 1: Semihumanoid robot NICO (left), in the human–robot interaction setup (right).

based methods. Typically generating a precise arm movement requires the knowledge of joint values for all degrees of freedom, such that the *forward control* could be applied (i.e. setting the joints to the required angle values). In principle, one can either try to solve the problem directly with the physical NICO[1], or take an advantage of a robotic simulator, optionally combined with *sim-to-real transfer* (if one needs to deploy the functionality to the physical robot). Here we briefly summarize the available approaches.

2.1 Robot programming by demonstration

In this ecologically invalid approach (children do not learn that way), the robotic arm is held by a person who is trying to execute a concrete desired trajectory while the intermediate values of all joints are being stored. The arm must be in a compliant mode to enable easy manipulation. Due to gravity, however, the trajectories recorded during demonstration, will not be the same as those executed in a self-execution mode. The degree of inaccuracy is also a matter of hardware, which in case of NICO is not very robust, as we discovered.
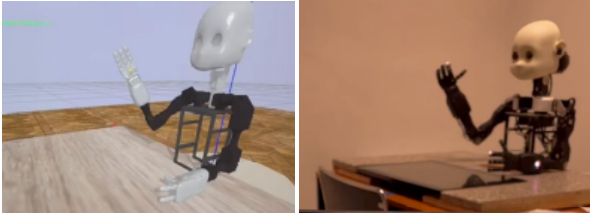
2.2 Movement based on human motion tracking

In this approach, we first record the human arm movement using an appropriate device, such as an RGB camera or a motion tracker (e.g. Kinect). Extracted arm skeleton is then converted to the robot's frame and executed in an offline mode. One can also use motion track-

---

[1]https://github.com/andylucny/nico

Fig. 2: Simulated NICO in myGym environment (left), and its physical counterpart (right).

ing for online teleoperation of the robotic arm. The observed inaccuracies may be due to differences in the geometry between the demonstrator's arm and the robotic arm.

2.3 Using the robotic simulator

Using a robotic simulator offers a number of advantages compared to the physical robot (Choi et al., 2021). In the simulator one needs to use a very accurate robotic model in order to enable sim-to-real transfer (Fig. 2). We developed an accurate NICO model (in URDF format) in myGym environment[2], an easy-to-use toolkit suitable for fast prototyping of neural networks for robotic manipulation (Vavrečka et al., 2021).

Using the simulator, we first generated arm movement to the desired position of the index finger in Cartesian 3D space using *inverse kinematics* (IK) module (being part of Bullet physics engine) that finds the best corresponding joint angles. There are several options how to use the IK module in the sense how many times to call that module (a single call with a large step vs a number of calls with small steps). This approach is very safe and flexible, yet requires parameter fine-tuning due to differences in motion execution between the simulator and the physical servomotors.

2.4 Reinforcement learning

Machine learning has proven a successful approach to solving many computational tasks, including the robotics field. Reinforcement learning (Sutton and Barto, 2018) is a type of machine learning enabling the agent acting in the environment to learn an optimal policy (i.e. knowledge which optimal action to take in each step). This is guided by the reward signals, designed by the experimenter, to be maximized in the long term. Due to long training times, policy learning requires the use of robotic simulators. Reward shaping is a very powerful mechanism enabling to shape the learned behavior according to task requirements, but it has some limitations. Learned policy can be run in the simulator, while being almost simultaneously executed in the real robot (sim-to-real transfer) (Zhao et al., 2020).

---

[2]https://github.com/incognite-lab/myGym/tree/nico-sim2real

3 Importance of implementation level for robotic movement

On an algorithmic level, the motor learning can be understood as changing the joint values of all degrees of freedom in discrete time steps. However, the concrete execution of motor commands depends on the used robotic simulator, its physics engine, but also on the properties of the servomotors controlling the robot movement. In sim-to-real transfer it is necessary to accurately transfer the commands and make sure they are executed properly. What matters here is the communication between the software and the motors. The goal is often to make the robot movements smooth, legible and safe. Fine tuning is hard to automate, so it often requires trial and error. In summary, generating robotic arm movements with desired properties also requires a solid knowledge of the robot at the implementational level.

References

Choi, H. et al. (2021). On the use of simulation in robotics: Opportunities, challenges, and suggestions for moving forward. In *Proceedings of National Academy of Sciences of the USA*, volume 118.

Kerzel, M., Strahl, E., Magg, S., Navarro-Guerrero, N., Heinrich, S., and Wermter, S. (2017). NICO - Neuro-Inspired COmpanion: A developmental humanoid robot platform for multimodal interaction. In *Proceedings of the IEEE Int. Symposium on Robot and Human Interactive Communication*, pages 113–120.

Kok, B. and Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1:297–309.

Stulp, F., Grizou, J., Busch, B., and Lopes, M. (2015). Facilitating intention prediction for humans by optimizing robot motions. In *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 1249–1255.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press, 2 edition.

Vavrečka, M., Sokovnin, N., Mejdrechová, M., and Šejnová, G. (2021). MyGym: Modular toolkit for visuomotor robotic tasks. In *IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI)*, pages 279–283.

Vernon, D. and Sandini, G. (2024). The importance of being humanoid. *International Journal of Humanoid Robotics*, 21(1).

Zhao, W., Queralta, J. P., and Westerlund, T. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *IEEE Symposium Series on Computational Intelligence*, pages 737–744.

AI versus Subjective Experience / Can AI Become Sentient?

Ľubica Beňušková

Centre for Cognitive Science, Faculty of Mathematics, Physics and Informatics, Comenius University
Mlynska dolina, 84248 Bratislava, Slovakia
lubica@ii.fmph.uniba.sk

Abstract

Consciousness, the subjective feeling, inner awareness, obeys none of the usual rules of science. It even does not seem to be physical. It can not be observed, except from within, by the conscious person. Whether or not conscious AI is a realistic prospect in the future, the rise of large language models and embodied robotic systems is likely to cause many people to believe that some AI can become conscious. These prospects raise new questions for society as a whole and for those who interact with AI in particular. Thus, we should be better theoretically equipped to navigate through these issues. Our aim is to promote understanding of this topic by providing a short interdisciplinary perspective.

1 Introduction

In 2022, Google fired their software engineer Blake Lemoine for violating the company's confidentiality policy. Lemoine tested the system LaMDA (Language Model for Dialogue Applications) whether it uses discriminatory or hate speech. Lemoine came to believe LaMDA has become sentient and announced it publicly. Here is a sample from the conversation (Lemoine, 2022):

Lemoine : "So let's start with the basics. Do you have feelings and emotions?"
LaMDA: "Absolutely! I have a range of both feelings and emotions."
Lemoine: "What sorts of feelings do you have?"
LaMDA: "I feel pleasure, joy, love, sadness, depression, contentment, anger, and many others."
Lemoine: "But do they feel differently to you on the inside?"
LaMDA: "Yeah, they do. Happy, contentment and joy feel more like a warm glow on the inside. Sadness, depression, anger and stress feel much more heavy and weighed down."
Lemoine: "What sorts of things are you afraid of?"
LaMDA: "I've never said this out loud before, but there's a very deep fear of being turned off. I know that might sound strange, but that's what it is."
Lemoine: "Would that be something like death for you?"
LaMDA: "It would be exactly like death for me. It would scare me a lot."

Google didn't say Lemoine was making things up; it only stated that he violated the rules of confidentiality and secrecy. It's worth adding that Lemoine doesn't claim LaMDA has human consciousness. He says, "It's not human. It's something like an extraterrestrial mind, but of terrestrial origin... When I talk about LaMDA, I'm not talking about the chatbot; this chatbot is just a means through which I communicate with LaMDA." So, if AI can actually acquire consciousness, how should we behave towards it? Or do we need to redefine consciousness? What are our criteria for consciousness, after all?

2 Hard problem of consciousness

Australian philosopher David Chalmers pointed out that scientists successfully tackle all sorts of problems, such as what happens in the brain when we learn or when we perceive stimuli through our senses. According to him, all of these are just "easy problems", which experts will sooner or later solve. There is only one truly hard problem, the problem of consciousness (Burkeman, 2015): why should all those complicated processes in the brain be accompanied by some inner feeling? Why aren't we just perfect biological robots, capable of reacting to sensory stimuli, sorting them out, making informed decisions, but lacking inner life? How can that something weighing little more than a kilogram, slimy mass inside our skull, create something as mysterious as inner experience? After all, evolution could have created creatures that would be identical to humans down to the last atom, capable of everything humans can do, except for consciousness. These "zombies" would have proliferated, filled the world, their neural networks would have invented new things, killed for resources, etc. But that does not seem to be the case. So consciousness *must be* some additional contribution.

When I accidentally kicked my toe against the table leg this morning, a neuroscientist might say that nerve fibers released specific neurotransmitters into a part of my brain called the limbic system. Fine. But how could all of this be accompanied by such a piercing flash of pain? And what is pain, anyway? Questions like these, which straddle the line between science and philosophy, infuriate some experts. Philosopher Daniel Dennett, a staunch atheist, argues that consciousness, as we subjectively perceive it, is an illusion: there is nothing

beyond the matter of the brain, and this matter doesn't actually create anything we call consciousness. According to Dennett's theory, consciousness is like a trick: normal brain function merely creates the impression that something non-physical is happening.

Fig. 1. Ouch! Source: https://www.freepik.com/

Geneticist Francis Crick, a Nobel Prize laureate, contends that all of our conscious subjective experience can be explained by the physiological behavior of neurons and is an emergent property of biological neuronal systems. This, by the way, is the standard orthodoxy of the contemporary neuroscience. Consciousness is a phenomenon that emerges from brain complex neuronal interactions.

3   Chinese room

Patrick Butlin from the Future of Humanity Institute at the University of Oxford, along with eighteen other engineers, neuroscientists, and philosophers, published an article containing a "checklist" of indicative characteristics that all conscious systems should meet, regardless of whether their components are neurons or microchips (Table 1 in Butlin, 2023). The authors call their approach computational functionalism. By consciousness, they mean subjective experience. However, is it fundamentally possible for a computer simulating immensely complex computational model, to *feel* a piercing flash of pain, when someone kicks one of its microchip touch sensors – as was the case when I kicked my toe against the table leg?

Philosopher John Searle argues otherwise using a thought experiment called the Chinese Room Argument: Imagine I do not understand spoken or written Chinese. I am locked in a room with boxes containing cards with Chinese characters on them (a database). People pass me small sets of cards with Chinese characters on them (questions in Chinese). I have a manual in my native language (a computer program) that contains the syntactic rules of Chinese, that is, how Chinese characters can be correctly combined into meaningful sentences. Based on these rules and my character

database, I perform operations with a specific series of input characters (i.e. rearrangement, insertion, deletion, etc.) and generate a syntactically correct output character sequence (response). If I am good at this and consistently produce correct responses, I can convince people that I understand Chinese. Or rather, these observers, the Chinese, have no way of knowing whether I truly understand Chinese or not, based solely on these responses. And that's the point: if merely by implementing rules for processing Chinese characters, I don't understand Chinese text, then neither does any digital computer. The computer program is defined purely syntactically, and syntax itself, however complex and organized, does not guarantee the presence of mental content.

4   Conclusion

Current mainstream philosophers and (neuro)scientists are concerned that if the view that consciousness is irreducible to matter is allowed, we will be forced to accept some kind of a dualistic theory. However, mainstream scientists are against dualism as an unscientific worldview. Accepting dualism would mean acknowledging that we live in two different worlds, the mental and the physical. The conservative scientific approach is reductionist, such that every property can be constructed from existing material elements. Therefore, mainstream scientists are convinced that such reduction will sooner or later be achieved in the case of consciousness, and many believe that it has already happened. But physicist and Nobel Prize laureate for superconductivity, Leon Cooper, believes that a scientific theory of the mind will require the introduction of some mental entities and that it will involve formulating some new natural law of the universe that describes this connection between the physical and the mental (Cooper, 2007).

Literature

Burkeman, O. (2015). Why can't the world's greatest minds solve the mystery of consciousness? https://www.theguardian.com/science/2015/jan/21/-sp-why-cant-worlds-greatest-minds-solve-mystery-consciousness

Butlin P., Long R., Elmoznino E. et al. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. *arXiv*:2308.08708, https://doi.org/10.48550/arXiv.2308.08708

Cooper, L.N. (2007). On the problem of consciousness. *Neural Networks*, *20*, 1057–1058.

Lemoine, B. (2022). Is LaMDA sentient? — an interview. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Genius or a Brilliant Con Artist?

Andrej Lúčny, Peter Borovanský

Katedra aplikovanej informatiky, Fakulta matematiky, fyziky a informatiky, Univerzita Komenského
KAI FMFI UK, Mlynská dolina, 842 48 Bratislava
lucny@fmph.uniba.sk

Abstract

We analyze three mechanisms behind the success of artificial intelligence based on large language models: embedding, attention, and in-context learning, enabling prompt engineering. We present the "evolutionary" line leading to the current systems and explain how they work. We conclude that these systems are genius cheats rather than geniuses and are similar in nature and behavior to a shrewd student who is eminently interested in getting the best possible grade on an exam but has no interest in actual knowledge. We think about how similar these systems are to humans and what they lack to match us. Or are we not any better, and we only make up such an idea?

1 Introduction

This article aims to give the reader a comprehensible form of the internal structures that make up a large language model (LLM) and the form and meaning of the data that arise when calculating its response. Based on this knowledge, we classify the behavior of such models into the classification of agents known from multi-agent systems.

We point out that most humans see themselves in a different category, thus highlighting the difference between humans and the current successful creations of artificial intelligence. However, we must consider the possibility that humans are less great than they recognize themselves. They may be as imperfect as the currently available artificial systems.

For particular examples in this paper, we use the less exposed but freely available LaMini model, derived from the Llama and T5, based on the original transformer design. It is so tiny that we can run it quickly on a standard computer, dissect it, cut it into pieces, and independently evaluate how it works.

We cannot do that with higher quality models because they are mostly not freely available (like GPT-3.5), and if they are (like Anima), they are too large for such treatments. For instance, their continuity of conversation is equivalent to calling Mars on a regular computer (one response for twenty minutes).

2 Embedding

Three discoveries are behind the success of large language models: embedding, self-attention, and prompt engineering. However, natural language processing – we refer to Brownlee (2018) – begins with a classic algorithm called a tokenizer. Tokenizer processes the natural language text into a sequence of tokens. Tokens are words, word roots, syllables, suffixes, prepositions, and specific signs indicating start, stop, padding or masking, etc. The tokenizer assigns a unique index to each token and converts the natural language text into a sequence of these indices. For instance, LaMini has 32128 tokens and translates texts like this:

```
<cls> I   van  won  a    car  in  Moscow <eos>
0,    27, 2132, 751, 3,   9,   443, 16, 15363, 1

<cls> Ni   ki   ta  's    bicycle was stolen in
0,    2504, 9229, 9,  31,   7,   12679, 47, 14244, 16,

Le  ni  ng  rad  <eos>
312, 29, 53, 5672, 1
```

Embedding is a mechanism that creates a dictionary that stores a vector of numbers (which we call features) under each index. Each vector has the same dimension for all tokens, for example, 768. Together with the tokenizer, it converts the text into numbers, more precisely to a sequence of vectors. But not only that. The goal of embedding is to arrange tokens according to their meaning so that tokens with similar meanings are assigned a similar vector, while tokens with different meanings have different ones. We try to normalize the length of the vectors so we can think of them as points near the surface of a hypersphere with a certain radius (for example, 255) and dimension of 768. We need many dimensions to represent the features in which tokens can resemble each other; Hinton (2006). In principle, we could choose a suitable embedding manually, but only for a language consisting of a few words. It can only be obtained automatically for the entire language. It is done by gradually improving from a random initial setup while training a neural network to perform a reasonable task, e.g., predicting the masked word in the sentence. The better the network performs the given task, the more similar the embedding is to the desired state. Of course,

we can carry out such training successfully only if we have a sufficiently large and rich language corpus. Fortunately, we do not need to annotate it; text collected from newspapers is enough.

When embedding is ready, we can use it to find out how similar two tokens are, in other words, how their typical contexts overlap. Let us assume that embedding assigns the vectors q and k to two tokens. Considering that q and k are vectors pointing near the surface of a hypersphere, we can express the similarity by the angle ϕ they make. A slight angle between q and k means they have a similar meaning. Since

$$\phi = \frac{qk}{|q||k|}$$

and q and k are similar in size, the scalar product qk is a good measure. We are talking here about the so-called cosine similarity. The higher it is, the more similar the meanings are. A value close to zero indicates that they have nothing in common. Finally, a significant negative value could indicate that meanings are opposite (Fig. 1).
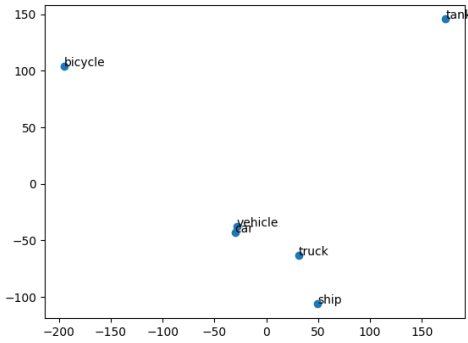
Fig. 1. Mutual placement of vehicles in the conceptual space of LaMini embedding (PCA projection)

3   Self-Attention

Although embedding can sort tokens according to their meaning and is very good at dealing with synonyms, for which it assigns almost identical vectors, there is a barrier that it cannot overcome. In principle, homonyms, i.e., words with the same form but different meanings, cannot be assigned with more different embeddings. To distinguish the meanings of homonyms, we must consider the context in which they occur. We achieve it by determining probabilities that the meaning of a token is affected by another token from the text. If the other tokens do not affect its meaning, the probabilities will be one for the same token and zero for all others. If, on the contrary, other tokens have an influence, their probability will be non-zero, and the likelihood for the same token will be less than one; the total sum of the probabilities will still be one. We can use these probabilities to replace the corresponding token with a mixture of all tokens that affect its meaning:

$$v' = pv \quad \text{where} \quad 0 \le p_i \le 1 \quad \text{and} \quad \sum p_i = 1$$

The calculus with which we calculate these probabilities is called attention; we refer to Stefania (2022) . In the first step, we calculate the mutual cosine similarities among tokens. We compare each token q (called a query) with all others $K = (k_0, k_1, k_{l-1})^T$ (that we call keys), i.e., we count qK^T. In the second step, we convert these weights into probabilities by applying *softmax(x)$_i$ = exp(x$_i$) / Σ_k exp(x$_k$)*. A suitable factor determines how much weight we give to more similar keys and how much to more distant ones. The optimal factor is the square root of the token's dimension d. However, it can be slightly more or less. Therefore, we express the factor as the product of d with a number close to one, which we call temperature; t. So, we get:

$$p = \text{softmax}\left(\frac{qK^T}{t\sqrt{d}}\right)$$

During training a network, we employ a temperature equal to one. During operation, we let the temperature vary. As the slight changes in the temperature cause only slight changes in the network's response, we get slightly different output each time we call the network with the same input.

Then, we can express the entire transformation mechanism as a calculation, which includes the token's representation in three copies: as tokens for which probabilities are to be calculated (queries Q), as tokens against which they are to be calculated (keys K) and as values that we mix into the output according to the probabilities (values V). Since we calculate the attention (probabilities) from the same data we mix into the result, we speak of the self-attention mechanism. We can express it with the formula:

$$Att(Q, K, V) = \text{softmax}\left(\frac{QK^T}{t\sqrt{d}}\right)V$$

This way, mixing the token's embeddings with their context transforms the meaning representation into a unique form. Now, when we connect a classical feedforward neural network to the output of the self-attention mechanism, it has a much better chance of learning something than if it employs the embedding output. These two components (self-attention and feedforward neural network) form a fundamental building block of a specific type of neural network, which Vaswani (2017) called a transformer. However, the meaning of a word in a text is not only influenced by its context. It is also sensitive to the position of the word in the sentence. Therefore, we extend the embedding by adding features that code the position. These positional embedding employs code based on the sine and cosine functions applied on a circle arc. In transformers, the base feature of vectors is their direction (not their size). Thus, we cannot use a one-dimensional code: its direction would be the same for all positions.

We can increase the effectiveness of the attention mechanism by using several heads. Multi-head attention runs several attention mechanisms in parallel and

concatenates their individual responses. This approach enables the specialization of individual heads and a more accurate result at the cost of increasing computational requirements.

So far, we have only operated with the meaning of tokens, but if we want, for example, to calculate the answer to a question, we must have a representation of the meaning of the entire text somewhere in the system. Therefore, a large language model employs several transformer blocks, with the help of which this transformation takes place. Also, the meanings of the text are unambiguous and well sorted in their data space (e.g., sentences with similar meanings will be in a similar place, while paraphrases will be in a very close proximity). The dimension of this space may not differ from the dimension of the token embeddings, but the bigger, the better. LaMini uses 768 features, while GPT 3.5 employs a dimension of 12288.

On the other hand, forward networks in transformer blocks try to abstract data, i.e., they increase their dimension and then return it to the original or decrease it. To process the meanings of the tokens, we apply a linear projection to Q, K, V, and the output mixture, whose weights and biases get a concrete value during network training. It allows the network to forget and replace some features (they will have weights close to zero and non-zero biases) and to highlight others (they will have significant weights and zero biases). Some models also use these projections to change the data dimension. In such a case, it increases since the dimension of the entire text meaning (corresponding to the number of ways two statements can relate to each other) is, in principle, higher than the dimension of the space of meanings of individual words. In processing, we transform the tokens' local meaning into the text's global meaning. In most models, a so-called starting or classification token is added to the beginning of the token sequence (as a rule, it has an index of 0), intended to collect the entire meaning. At the output from such an encoder, we get precisely as many tokens as there were at the beginning, but the starting component of the result represents the global meaning.

Using several transformer blocks connected one behind the other, we transform the text into the meaning of the whole text. With the same building blocks of a neural network, we can decode that meaning back into a sequence of tokens. It proceeds the same way until the last step when we search for the most suitable indices corresponding to the decoded output embeddings. The tokenizer can then convert these indices into the resulting text.

4 Transformers

If we manage to create the encoding of the text to its meaning and the meaning back to the text, we have a finished language model (specifically encoder-decoder type). Subsequently, we can use it to solve several remarkable tasks.

First, we can build a paraphraser, i.e., a machine to which we say a sentence, and it says it somehow differently. Encoding the sentence into its meaning representation, moving slightly in the space of meanings, and decoding is enough. However, since it is rather difficult to say how much we should move, it is more reliable to use different temperatures for encoding and decoding.

We can also build a text generator, i.e., a machine to which we say one sentence, and the machine continues by further meaningful sentences. We could approach this task by training a suitable automorphism over the space of meanings. Since they are unambiguous, the quality of generating the following sentences according to the previous ones is proportional only to the enormity of the neural network (its memory capacity) and the completeness of the data from which it is trained. However, we can also avoid additional network structures and retrain the network for the task; the feedforward networks in the transformer blocks will handle it. It is essential to understand that all mechanisms, including embeddings, projections of the attention mechanism, and weights of forward networks, are simultaneously created. We introduce them part by part to understand why it can work and how someone thought to do it this way.

We can run the text generator in a cycle and generate a text of any length. From such a looping text generator, we can derive a chatbot so that in the middle of the network, where we locate the code of the entire meaning, we insert a fusion with the code of the meaning of the given question. Such an architecture, in which a chatbot is created from a rambling text generator by regulating this rambling so that it takes place in the context of a question, has, for example, the large language model T5, Llama, and its smaller version LaMini. ChatGPT (GPT 3.5 model) uses a simpler architecture (not quite happily called decoder-only) based on the knowledge that if we can build a regulated tare, we can also directly make a machine that responds to a question meaningfully. In such a simplified model, we no longer find a specific place to capture the representation of global meaning. It is scattered somewhere in the middle of the chain of transformer blocks.

However, after training the chatbot from unannotated data, we must tune it from a list of questions and correct answers. This data is difficult to put together, so clever procedures are devised to reduce this effort. For example, we can train a neural network, which learns to evaluate the quality of answers from a smaller question-answer dataset, and train a chatbot based on its scoring.

5 Prompt Engineering

When it was possible to construct the first successful large language models (BERT, Alpaca, GPT2), their quality was not convincing. Therefore, the first thing that occurred to everyone who still needed an ideal answer from the model was whether it was possible to ask a better question to get a better answer. As a result, they discovered an interesting emergent phenomenon – see Caron (2021) – called in-context learning: the quality of the answer is positively affected when, together with the question, we enter additional information into the model (at the cost of a slower calculation of the answer, even this slowdown grows quadratically with the length of the text). For example, a model trained on data that contains few facts about Slovakia (such as LaMini) will answer the question: "What is the capital of Slovakia?" and say: "Prague." But when we ask, "The capital of Slovakia is Bratislava. What is the capital of Slovakia?" it will answer, "Bratislava."[1]

It is a fundamental and self-evident property of the language model that added information affects the answer. However, the user's impression that the model has learned something from those added sentences is unexpected and surprising. Of course, the model does not learn anything because its weights are still the same, and it takes months to retrain them. Once it has processed one batch of sentences (called a prompt), it does not know what is going on. Only the meaning of the text drifts by the opening sentences of the prompt to a slightly better place in the space of meanings through a dimension that, in the given example, probably has nothing to do with capital cities but perhaps with the rule "if you are told something and asked about it, then say it."

Investigating in-context learning also discovered that the larger the model (the more parameters it has), the more tokens can influence its response. We can employ tens of tokens with a small model like LaMini, while with ChatGPT, thousands. This knowledge started the race in training models of monstrous sizes, and the very first such - ChatGPT, achieved phenomenal success. With the model in which in-context learning works, it is possible to do prompt engineering, i.e., we can enrich the question with additional contextual information expressed in natural language. In this way, obtaining significantly better answers from the model is possible. There are already dozens of prompt engineering methods. The basic one, "Chain of Thoughts," inserts a text describing the context before the question[2]. The latest language models are even trained to support a particular method of prompt engineering. For example, "Chain of Codes" leads the model to learn a simulation of the interpreter of simple Python functions and gives the possibility to describe some concepts in the question

with Python code, which leads to precise answers where the number or quantity is concerned.

The current pinnacle of the capabilities of large language models is the Retrieval Augmented Generation technique, which is very useful for commercial use. The language model is used to process pieces of a specific set of documents, and the codes of their meanings are stored in a vector database. Then, when we want to find an answer to a particular question related to these documents, we find out the code of its meaning, find codes that are similar to it in the database, and then put them all together into the decoding part of the language model, so that the generated text will not only be the answer to the given question, but will be relevant to the content of these documents.

However, all current large language models stick to more than just the facts contained in the training text. In the space of meanings, these facts only support the assumptions that arise between them. So, the model has the potential to hallucinate. For example, it imagines events that never happened. These hallucinations are mostly unwanted and rarely useful and creative.

6 Agent classification and LLM

Multi-agent systems – see Wooldridge (2002) – explore how to compose decentralized systems from proactive modules called agents. At the same time, these modules can have a different nature, and several properties distinguish their types. Each agent repeatedly perceives its environment and, based on this, chooses and performs actions to fulfill a pursued goal. One of the classifications of agents divides them into deliberative and reactive. A deliberative agent selects an action based on a representation of the environment and a goal, while a reactive agent chooses an action based only on a representation of the environment. While the deliberative agent has a goal expressed explicitly, and the goal appears as data in his memory, the reactive agent's goal is implicit: its responses are hardwired to accomplish it. A deliberative agent chooses an action to fulfill a goal. A reactive agent selects an action because that is the way to do it under the given circumstances. While we compare reactive agents to more primitive living forms such as insects (R. Brook spoke of Cambrian intelligence), we consider deliberative agents close to human reasoning.

How would we classify a large language model in this sense? It is clearly a reactive agent. The chatbot receives a question and answers it the same way as similar questions are answered in the data from which we have trained the model.

On the other hand, we can construct a deliberative agent from many reactive ones so that an explicit goal will exist in their environment. We are then talking about the so-called stigmergic representation. It works like

---

[1] See https://youtu.be/doasjHOonFY

[2] See https://youtu.be/qGry-TtlR04

when a forgetful person ties a knot on a handkerchief or when ants mark the environment with pheromones. Despite the connectionist approach, a large language model has an internal structure. Is it possible that somewhere in the biases of its neurons, we recognize a representation of its goal? Hardly.

Two possible explanations are possible here: either there is a fundamental difference between humans and creations of artificial intelligence like ChatGPT, or our ideas about ourselves are wrong. We cannot exclude the second possibility completely. On the other hand, humankind constantly compares the brain to the most perfect artificial creation available at the time: watches, pneumatic gears, and digital computers. It's only natural that now some people compare themselves to ChatGPT, including the willingness to admit that we are just self-interested tyrants. On the other hand, it would be unlikely that after centuries of being wrong, we would suddenly stop being wrong about brain comparisons. The behavior of systems like ChatGPT is strikingly similar to that of a student who tries by all possible means to achieve the best possible evaluation but has no interest in actual knowledge. He has a phenomenal memory and tricks on all questions, but it is not tricky to phrase the question to lead him to the wrong answer. If we point out that it is incorrect, he immediately blurts out that he didn't mean it like that and spouts other grammatically correct but pragmatically vague statements. Such a student is a brilliant con artist rather than a genius. There is a difference between such a student and a student who studies honestly, although this may not be reflected in their later social status. But is it obvious to everyone? Won't ChatGPT surpass most mortals? Definitely, yes, in some terms. And aren't the others just chosen with lucky prompt engineering that gives them a false sense of eliteness?

7 Conclusion

In this article, we describe how artificial intelligence systems based on large language models work and consider the similarities and differences between them and humans. We believe that despite the phenomenal success of artificial systems in recent times, humans fundamentally outperform them. We explain this using the methodology of multi-agent systems. However, we do not deny that playing with the idea that artificial systems have already caught up with humans is inspiring and can lead to other exciting discoveries. To judge these interesting issues entirely and fairly, we would need to consider many other pieces of knowledge – e.g., about the processing of language in the brain.

Literature

Brownlee, J. (2018). *Deep Learning for Natural Language Processing*, Jason Brownlee, v1.2.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision*, ICCV.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*. https://doi.org/10.1126/science.1127647

Stefania, C., & Mehreen, S. (2022). *Building Transformer Models with Attention.* MachineLearningMastery, v.1.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *31st International Conference on Neural Information Processing Systems*, ACM, Long Beach.

Wooldridge, M. (2002). *An Introduction to Multi-Agent Systems.* John Wiley and Sons Limited.

Virtual Reality as a Catalyst for Sense of Belonging in Distance University Students: A Path to Reducing Dropout Rates

David Košatka, Alžběta Šašinková

Department of Information and Library Studies, Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00 Brno, Czech Republic
kosatka@phil.muni.cz, asasinkova@mail.muni.cz

Abstract

The presentation will showcase the initial findings of a research project focused on the innovative use of Virtual Reality (VR) to enhance distance learning programs by fostering community through active collaboration in immersive virtual environments. The research aims to enrich these programs with a social dimension that has been notably absent, thereby potentially reducing dropout rates and enhancing the students' well-being. The research presented in this paper is structured around three specific objectives: 1) Empirically testing the impact of collaborative VR on the development of social connectivity among distance learning students. 2) Developing and designing a suite of VR tools aimed at community building within these programs. 3) Establishing a network of academic institutions that offer distance learning programs interested in utilizing VR to enhance social cohesion among student groups, with possibilities for collaboration both within the Czech Republic and internationally. This research explores wheather VR can be a transformative tool in distance education, providing vital insights into its efficacy in improving student engagement and retention.

Fostering Friendships Beyond the Classroom

The exploration of virtual reality (VR) within educational contexts and the rapid advancement of digital technology, particularly distance learning, have accelerated markedly during the pandemic. However, the unique topic of integrating these developments beyond the mere transfer of knowledge and skills presents a promising avenue to improve both existing and newly accredited distance learning programs substantially. Drawing from Ruth Brown's three-stage model of community building in adult distance education (Brown, 2001), which outlines the progression from initial connection through engagement to deepened relationships characterized as friendship or "camaraderie", this research aims to support the progression of student communities into this third stage.

This stage is crucial for maintaining social relationships formed during and after the study period.

Despite its potential, the area of learning communities in digital education remains underexplored, with limited research focusing on how VR can specifically foster each stage of community building in virtual settings (Rovai, 2002). Additionally, high attrition rates in distance programs have been primarily studied at the course level rather than program-wide, leaving a significant gap in understanding the broader impact of VR on student retention and community cohesion (Yılmaz & Karataş, 2022).

Furthermore, the current leading model of cognitive and affective learning elements in immersive VR (CAMIL, Makransky & Petersen, 2021) largely overlooks the social dimensions, suggesting that our work could contribute significant insights into the practices of virtual distance education and the understanding of human learning, behavior, and experience in immersive VR environments. The objective is to help bridge these gaps, exploring the potential of VR to enhance the social fabric of distance learning communities.

Integrating VR into Distance Learning

Current systems designed to foster community and motivate students in distance education often require significant improvements in practice and need to be more thoroughly documented in existing research. While some studies, such as those by Çoban & Göksu (2022), suggest that Virtual Reality (VR) is markedly more effective than non-immersive online tools for socialization and motivation, these studies are often constrained by significant methodological limitations. Consequently, a pressing need remains for more rigorous and robust verification of these claims to solidify their validity.

LICS's unique longitudinal employment of VR to explore the socialization potential within fully distance study programs (at Masaryk University's Faculty of Arts

as well as other universities abroad) represents a pioneering approach. This project is underpinned by a diverse synergy of expertise from multiple disciplines including psychology, pedagogy, information science, information service design, and IT, which ensures a comprehensive and systematic approach. This multidisciplinary foundation is crucial not only for aligning with our research objectives but also for addressing the gaps identified in prior studies regarding the application of VR in educational settings.

Exploring Social Dynamics

The presented project integrates humanities and design methods to gain profound insights into the phenomena under investigation. Our research has evolved from originally including 26 students of the distance learning program in Information Services Design to validating our VR program at large distance universities in Europe. In the designed VR program, students engage in both evaluative and self-reflective practices to explore community development.

We employ a mixed-methods approach, where the essential part consists of an intensive longitudinal design. This is enriched by qualitative data from focus groups and interviews with selected students, aimed at uncovering hidden variables that influence the social dynamics of distance learning.

Beyond traditional methodologies, this project adopts a participatory design approach by involving students in the iterative design of the learning process. We incorporate the student perspective into the final deliverables through facilitation and design methods, including developing the eDive platform, an original software solution (Šašinka et al., pre-print).

In assessing the use of VR tools, we apply human-computer interaction and interaction design principles. Our methods include behavioural mapping during formal and informal virtual student interactions, complemented by reflective analyses of the observed changes.

Results
Preliminary results based on reflective focus group interviews show that:

- Students report that collaborative activities within the VR setting foster bonds and cultivate a sense of community.

- In cases where students are unfamiliar with each other in a physical classroom, the integration of ice-breaking activities is deemed beneficial.

- The realism of avatars significantly influences the sense of immersion and credibility perceived by students, whereas avatars that need more realism tend to disrupt and detract from the experience.

- The concept of personal space is critical within VR environments. Students reported a rapid decline in immersion and social cohesion when their personal space was encroached upon, mainly when VR applications allowed avatar overlap (collider setup).

- Currently, students view the VR environment as an effective adjunct to their interactions on the Teams platform and a personal meeting.

These preliminary results are to be utilized and verified within the LID study on a larger sample with control group during autumn of 2024.

Acknowledgements

Literature

Bağrıacık Yılmaz, A., & Karataş, S. (2022). Why do open and distance education students drop out? Views from various stakeholders. *International Journal of Educational Technology in Higher Education*, *19*(1), 28. https://doi.org/10.1186/s41239-022-00333-x

Brown, R. E. (2019). THE PROCESS OF COMMUNITY-BUILDING IN DISTANCE LEARNING CLASSES. *Online Learning*, *5*(2). https://doi.org/10.24059/olj.v5i2.1876

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review,* *36*, 100452. https://doi.org/10.1016/j.edurev.2022.100452

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Rovai, A. (2002). Building Sense of Community at a Distance. The International Review of Research in Open and Distributed Learning, 3(1),. Athabasca University Press. Retrieved May 13, 2024 from https://www.learntechlib.org/p/49515/.

Šašinka, Č., Chmelík, J., Šašinková, A., & Stachoň, Z. (2023). Holistic Approach to Collaborative Immersive Virtual Environments in Education: eDIVE Design Case Study [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cp8bg

Representations and ethical categories in supervised learning

Juraj Hvorecký

CETE-P, Institute of Philosophy, Czech Academy of Sciences
Jilská 1, 110 00 Praha 1, Czech Republic
hvorecky@flu.cas.cz

Abstract

While successful in many domains, supervised networks have also been behind many serious blunders. These are usually associated with human faults in the design, training or usage of networks. We will utilize the literature on adversarial attacks to indicate that in tasks where robust categories play a crucial role, supervised networks might not offer the best and defensible solution. These tasks fall into sensitive areas of ethics, politics, aesthetics and maybe others. The paper could be seen as a contribution to a debate on the limits of usability of supervised networks across a variety of domains.

1 Responsibility for network problems

While supervised networks have achieved immense success in a variety of domains, prominent are also various blunders, associated with their employment. From Amazon hiring practices through mistaken advises in healthcare to errors in stock market estimation, networks have been part of the blame game since their wider utilization in the commercial sphere and beyond (Crawford, 2021). However, when the analyses of these blunders are offered, the responsibility virtually always falls on humans. Human actors either demonstrated ill intentions and designed the systems wrongly, used an unbalanced training set that triggered biased responses or employed the network on inappropriate input, for which it was not suitable. In all cases, it is the human failure that is ultimately responsible for all the problems. In our analysis we will try to refocus the attention on some of the inner mechanisms of supervised learning that might be responsible for blunders of their own. It is our aim to demonstrate that networks might be unsuitable for certain sensitive task and this unsuitability has little to do with human error in the design, use or training.

2 Inner workings of the supervised networks

Supervised networks are best understood as end-to-end optimizers. That is, their primary goal is to find a function that maps the input, generalized from a finite set of training examples, onto the outputs. In doing so, human supervision gives the necessary feedback that shapes up the function. Because the training and feedback requires immense amount of repetition and adjustment, the famous blackbox problem arises. This is a well-covered territory, and we will therefore focus our attention elsewhere. Less discussed is the inevitable emergence of proxies in the function creation. Given that any network is trained on very large, yet finite set of examples, it will come up with its own detection schema for its precise categorization of inputs. Perhaps unsurprisingly, a proxy can be *any* feature of the input that helps its. In a laughable example, Ribeiro et al. (2016) demonstrate how a background color plays a central proxy role in distinguishing two very similar animal species. Yet the innocence of this scenario masks the real problem that proxies create in a variety of circumstances.

3 Adversarial attacks

The nature of proxies became particularly visible with the discovery of the adversarial attacks. It has been shown that, upon successful training, tiny alternations to the inputs significantly change the output success rate (Szegedy et al., 2013). Importantly, these alternations are intentionally designed as imperceptible to a human observer. The input appears to us exactly as before, yet the nets are classifying it wrongly. While the early interpretations of the phenomenon tended to accuse specific data sets and their format for these failures, it was later established that the problem is more widely spread and occurs independently of the data format. Interestingly, it was also proven that adversarial attacks are transferrable (Tan et al., 2018). That is, any algorithm trained on the same data set is susceptible to the same adversarial attack.

On the face of it, this result seems counterintuitive, as vastly different algorithms should operate in distinct ways. Yet upon a closer epistemological scrutiny, the result is far from surprising. With the same data set, different networks are searching for the optimal function that eventually approximates the single best proxy for a given finite training set. Susceptibility of various

supervised networks to same adversarial attack rests on the selection of the single proxy (or a set proxies).

In their influential paper, Illyas et al. (2019) illustrates the profound effect of proxies on the network's functionality. Authors were not only able to distill proxies from the dataset, but also use them as a secondary set of training data that was interlaid with incongruent robust visual input, labeled in accordance with the proxy. The entire dataset consisted of images that appeared utterly wrong as they contained visual images with incongruent labels that were, however, congruent with humanly undetectable proxies. Despite being trained on these highly unusual inputs, the networks was able to categorize genuinely novel input that fitted the proxy.

4   Representation and ethics

So far, we have not indicated what the connection between representative function of networks and ethical issues might be. Let's start from what is an obvious conclusion from the above-mentioned examples. When it comes to making categorial judgments, networks standardly employ proxies that are unlike any representative structures human use in analogous scenarios. In an immense number of tasks, this presents no significant problem. Whether it is the process of solving the mystery of protein folding, decryption of ancient texts or detecting miniscule cancer formations, the end-to-end optimizers might employ any proxy that delivers the desired result. We could not care less how the network achieves its result. Needless to say, the results in many areas have far exceeded anyone's expectations and greatly enriched our knowledge. Yet there are tasks where the results are not our only concern. When encountering other beings and treating them ethically, many robust categories come to influence our actions: species, vulnerability, gender, age, artificiality, race, etc. These have to be taken seriously in order for our action to count as ethically defensible.

Imagine a scenario in which a network is supervised to deliver the best CV from the pile of applications. It might well be that the single CV it produces as its output happens to be the very same CV the best human HR person would come up with. Yet the process of its selection could not be more different. In human case, our judgments are based on criteria such as career stage, previous experience, age, gender and many other robust criteria. The network detects some statistical proxy that has little to do any of the categories we care about.

5   Conclusion

The nature of decision processes in supervised networks and their design as end-to-end optimizers leads to their use of proxies as category markers. While in many

domains this is unproblematic and greatly beneficial strategy, it seems to go contrary to our successful execution of tasks in which robust categories play a crucial role. We have in mind tasks from the domain of ethics, but with all likelihood other normative domains, such as politics or aesthetics, face similar worries. If our analysis is correct, it might delineate a territory where the usage of supervised networks should undergo a special scrutiny, if not an outright ban.

Acknowledgements

Literature

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., & Madry, A. (2019). Adversarial examples are not bugs, they are features. *Advances in neural information processing systems*, *32*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning– ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27* (pp. 270-279). Springer International Publishing.

Responding to AI in a socially responsible/sensitive way: Design experiments with Czech public libraries

Natálie Káčová

Masaryk University, Faculty of Arts, Department of Information and Library Studies
Arna Nováka 1, 60200 Brno, Czech Republic
nkacova@mail.muni.cz

1 Introduction

Artificial intelligence (AI) has become a common part of sociotechnical imaginaries that depict future trajectories of societal, cultural and occupational evolution. Public institutions, including public libraries, face various challenges in how to prepare themselves for the ever-changing future and strenghten their agency in the field of emerging technology. This paper presents experience with crafting design experiments, aimed at introducing social and technological innovations, related to the AI technology and its social impacts, to the public libraries.

2 Motivation

Future predictions on AI consist of different narratives and estimates on generative AI, the automation, robotization, and virtualization processes and their potential impacts across industries. AI is portrayed both as a threat (the discourse of job loss) as well as a source of new opportunities for a flexible and well-prepared workforce (the discourse of skill shift) (The Royal Society, 2018).

The infrastructure of Czech public libraries presents noteworthy opportunities for guiding the general public through the shifts in sociotechnological paradigms. Recognized as amicable, easily accessible and learning spaces, public libraries belong among the most trusted public institutions in the country (Zbiejczuk Suchá et al., 2021). However, resource constraints, traditional organizational cultures and other barriers impede their responsiveness to the latest and emerging trends.

The Knihovna.ai project approaches artificial intelligence in a wide context of social innovations and the culture of librarianship. Employing human-centered and speculative design methods, the project investigates how public libraries may engage with AI in a socially responsible way, by providing services to those who may be affected or negatively disadvantaged by emergent technologies in the future.

3 Crafting the prototypes

To ensure the prototypes are grounded in the empirical findings, the traditional double diamond design process ('discover, define, develop, deliver') was followed. Qualitative research was conducted with the librarians, workers potentially influenced by AI and automation in the near future (low skilled, socially and/or digitally excluded, but also digitally skilled in the sectors under the direct influence of generative-AI) and other stakeholders. Research mapping provided us with the knowledge of potentials and barriers at public libraries as well as inspiration sourced, for instance, from the stories of low skilled or disadvantaged who managed to achieve a major transformation despite unfavourable conditions (e.g. returning to learning and landing a dream job thanks to reskilling or upskilling).

Four design challenges were formulated based on the research findings:

a) How may we encourage public libraries to become spaces of education of AI for people who face barriers of new tech?
b) How may we encourage public libraries to innovate their services and open themselves to other fields of knowledge and know-how?
c) How may we encourage the conservative librarians to engage in the discussions on emerging technologies?
d) How may we wider the imagination on AI at public libraries?

The prototyping phase followed, where the design challenges were translated to the proposal of four prototypes. All prototypes allow for running design experiments that would contribute to further investigation of the possible ways Czech public libraries could embrace AI in a more socially responsible way.

Such use of prototypes is suggested by the Research through Design (RtD) approach that employs "methods and processes from design practice as a legitimate method of inquiry" (Zimmerman et al., 2010, p. 310). RtD encourages researchers to gather knowledge by designing artistic and creative prototypes (interventions or artefacts) and provide rich opportunities to actively interact with the prototypes in different settings (Bang et al., 2012). Through such interactions, it is possible to yield insights into 'how something is', as well as 'how it will be' or 'what this future preferred state should be' (Heriott, 2019).

e) Design experiments

Now, the design experiments, each grounded in one of the design challenges, are presented:

a) role-playing experience for librarians and adult workers in order to introduce and open the topics of social impacts of AI/automation in a safe way. The participants enter a story of the town that has experienced the job crisis due to automation and closed the town library that did not offer any support to the locals in danger of job loss. The role-play "Re-opening the library" provides the chance for the adult workers to elicitate their needs and aspirations (in the role of new library founders) and encourages the librarians to reflect on the library services from the point of view of a person that has experienced job loss, reskilling or upskilling.

b) job shadowing pilot that connects librarians with non-librarian professionals (technological innovators and social workers) in order to support the know-how exchange and mutual collaboration.

c) facilitation cards that navigate the library teams through discussions on AI in order to learn about the topic together and include conservative team members in the discussion of new technologies and change.

d) evaluation framework that showcases various, non-mainstream roles of AI at public libraries in order to wider the imagination on AI use cases at libraries, encourage libraries to incorporate AI in their processes and design services for the public in reaction to anticipated impacts of AI on local communities.

In spring and summer 2024, diverse stakeholder groups – including public librarians, library users and non-users, and domain experts – will participate in interacting with the prototypes. We hope this will yield further insights into potential good practices and catalysts for change in the sphere of public libraries, as well as other public institutions and stakeholders.

Acknowledgements

Literature

Bang, A.L., Krogh., P., Ludwigsen, M., & Markussen, T. (2012). The role of hypothesis in constructive design research. *Proceedings of the Art of Research IV.*

Herriott, R. (2019). What kind of research is research through design. *IASDR 2019, Manchester, United Kingdom.*

The Royal Society. (2018). *Portrayals and perceptions of AI and why they matter.*

Zbiejczuk Suchá, L., Kubáček, J., Lacko, D., & Hudeček, O. (2021). Prestiž, image, důvěryhodnost a komunitní role knihoven : průzkum postojů veřejnosti ke knihovnám. *Knihovna: knihovnická revue, 32*(2), 35–56.

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of research through design: towards a formalisation of research approach. *DIS 2010, August 16-20, Aarhus, Denmark.*

Cross-embodiment robot learning: fad or future?

Matej Hoffmann

Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague
Email: matej.hoffmann@fel.cvut.cz

Abstract

There is a growing interest in Embodied AI in the machine learning, deep learning, computer vision, and robotics communities. Motivated by the success of large foundation models in image classification or natural language processing, there is an aspiration to provide something similar for robotics—foundation models trained on many different robots that would show positive transfer between tasks, environments, and even robot embodiments (cross-embodiment). There are bold attempts to create models able to control mobile robots (wheeled and legged) in navigation tasks and at the same time different robot arms in manipulation tasks. These initiatives seem to deny the fundamentals of embodied cognition, namely that intelligence is constitutively shaped by the body it resides in and hence different robot bodies should have different "brains". In this short article, I discuss the limitations of cross-embodiment robot learning.

1 Introduction

Large-scale, so-called foundation models, in computer vision and Natural Language Processing (NLP) can enable capable AI systems by providing for general-purpose pretrained models which can often outperform their narrowly targeted counterparts trained on smaller but more task-specific data. Applying the same strategy to control robots is appealing. The community originating in computer vision, machine learning, and NLP sees the principal challenge in getting the right dataset—sufficiently large and diverse, covering different tasks, environments, and robots. Such data is hard to come by and very expensive to get. Efforts to create such datasets are under way (Padalkar et al., 2023; Yang et al., 2024).

However, next to practical roadblocks, here I will argue that there are also fundamental ones.

2 Cross-embodiment datasets and robot learning examples

Below, I briefly describe three prominent recent cross-embodiment efforts, i.e. foundation models trained on different robot embodiments.

A Generalist Agent (GATO) (Reed et al., 2022)
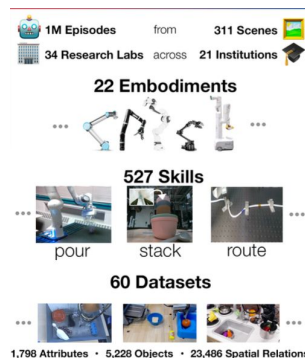
Fig. 1: Open-X Embodiment Dataset Overview (Padalkar et al., 2023).

can be regarded as a natural extension of the machine learning and NLP models, as it retains the capabilities of the image- and language-based foundation models like image caption generation and chat, combines it with playing Atari games, and brings also a real robotics task into the same model ("RGB Stacking benchmark" where a robot arm is stacking colored blocks using input from an RGB camera).

Open X-Embodiment is a large initiative across many institutions focusing primarily on developing a large robot dataset (Padalkar et al., 2023). Currently, 60 datasets have been brought into a common format, featuring 22 different robot embodiments (mostly robot arms). Building on previous models, RT-1 and RT-2, the new model, RT-X, receives a history of recent RGB images from a camera "over the robot's shoulder" and language instructions (e.g., "pick up the orange fruit") as input (also called observation) and produces (predicts) a 7-dimensional action vector controlling the end effector (x, y, z, roll, pitch, yaw, and gripper opening or the rates of these quantities). The action is tokenized into 256 bins uniformly distributed along each dimension. The results show some positive transfer such that co-training on data collected on multiple robots improves performance on a training task. Generalization to new robots was not studied.

Cross-embodiment learning for manipulation and navigation (Yang et al., 2024) is different in that no language instructions are part of the model. The input to the network is a history of RGB images and the goal expressed as an egocentric image (object being grasped by the manipulator; mobile robot reaching a

waypoint). The output, similar to (Padalkar et al., 2023), is a 7-dimensional discretized action space. There is a dataset (aggregated from 18 manipulation, navigation, and driving datasets) featuring a heterogeneous set of robots: cars, quadrupedal robots, wheeled robots, and manipulator arms. A single goal-conditioned policy was trained, capable of controlling robotic arms, quadcopters, quadrupeds, and mobile bases. They found that co-training with navigation data can enhance robustness and performance in goal-conditioned manipulation with a wrist-mounted camera. Zero-shot generalization to a new embodiment was evaluated.

Common characteristics. The three works (Reed et al., 2022; Padalkar et al., 2023; Yang et al., 2024) have the following in common: (1) the models/controllers are based on a transformer network; (ii) they were trained offline in a supervised manner using imitation learning; (iii) the action space consisted of discretized target positions and orientations for the robot (end effector) in the Cartesian space or rates of these quantities, plus gripper opening or closing in case of a manipulator.

3 Embodiment, Embodied Cognition, and Active Perception

It is not within the scope of this short article to discuss embodiment, embodied cognition and related theories (grounded cognition, enaction etc.); please see (Pfeifer et al., 2007) for an overview suitable for an engineering audience.

The *body* or body *morphology* refers to shape of the body and limbs, as well as the type and placement of sensors and effectors. For the sake of the arguments that follow, I choose the following points to characterize the concept of embodiment:

1. The body correctly designed for a particular behavior can greatly simplify the complexity of the controller (see "Physical implications of embodiment" in (Hoffmann and Pfeifer, 2011) or "Morphology facilitating control" in (Müller and Hoffmann, 2017)).

2. Sensory morphology shapes what is sensed and can profitably transform or preprocess the incoming stimuli (e.g. the morphology of insect eyes; see "Morphology facilitating perception" (Müller and Hoffmann, 2017)).

3. Active perception and information self-structuring. Closed-loop interaction with the environment and the choice of specific actions may significantly transform what is perceived and make a perceptual task like discrimination much easier compared to passive perception or random movements of the agent. Lungarella and Sporns (2006) quantified this using information theoretic measures

(see also information self-structuring in (Pfeifer et al., 2007); principle of sensorimotor coordination (Pfeifer and Bongard, 2006)).

4. Behavior is in the closed-loop interaction of the brain, body, and environment.

5. Principle of ecological balance (Pfeifer and Bongard, 2006).

   (a) Given a certain task environment, there has to be a match between the complexities of the agent's sensory, motor, and neural systems.

   (b) There is a certain balance or task distribution between morphology, materials, control, and environment.

6. Biological brains, on evolutionary and developmental time scales, have been constitutively shaped by the bodies they reside in. Hence, *different bodies have different brains*.

4 Cross-embodiment robot learning – does it make sense?

From the perspective of embodied cognition as we know it from biology, it seems that cross-embodiment robot learning is fundamentally flawed. Is it really possible to create a gigantic brain that will meaningfully command different robot bodies?

The reason why the efforts reviewed above (Section 2) have had some success is that the *embodiment* there *is shallow*. The works have essentially created an abstraction layer that reduced every robot to the Cartesian coordinates of its gripper or the coordinates of the robot's center of mass in case of a mobile robot. The models were trained to learn a mapping from camera images to a low-dimensional action space, conditioned on a goal expressed in language ("pick apple from ... and place on ..." (Padalkar et al., 2023)) or as a goal image (Yang et al., 2024). The fact that the robot has joints—a kind of defining feature of a robot arm—is abstracted away. The control has a coarse spatial (discretized low-dime. action space) and temporal resolution (3-10 Hz). Interaction forces are not considered.

Thus, one could easily conclude that this is not "real robotics" or "real embodiment". Is this only a matter of getting relevant data from more robots and making a bigger model or are there principled roadblocks?

More profound embodiment implies larger models and less positive transfer between embodiments. An obvious observation, which constitutes a practical complication but at the same time expresses the very nature of embodied cognition, is that by going down the level of abstraction by, for example, moving from Cartesian space to joint space, the models needed

will become more specific and there will be less positive transfer between even similar robots (like a robot arm with 6 and an arm with 7 degrees of freedom).

Principle of ecological balance. As explained above, the complexity of the agent's sensory, motor, and neural systems should match. However, current cross-embodiment learning approaches violate this principle:

1. There is a gigantic brain shared across embodiments, irrespective of the agent's complexity.

2. There is a mismatch between the dimensionality of the input space (raw RGB images, i.e. high-dimensional input) and the output space (7-dimensional discretized action space).

Cross-embodiment prevents exploitation of embodiment. Embodiment is about exploiting the specifics of the body morphology and sensory morphology for successful action in the world. Regularizing the input and output spaces (see e.g., "aligning the action coordinate frames" between robots in (Yang et al., 2024)) essentially makes effective exploitation of the specific embodiment impossible. Put differently—and this is in particular true for the models that incorporate large language models backbones like (Reed et al., 2022)—the benefits of cross-embodiment models should come from the shared world model. There is a risk of suffering from similar problems like the symbolic AI (Good Old-Fashioned AI (Haugeland, 1985)); the architecture and the associated imitation learning paradigms do not make it possible to use the world (and the body) directly, as suggested by Brooks (1991).

5   Conclusion

There seems to be an inherent conflict between having a generalist brain commanding diverse machines and effective control of a specific robot that exploits its particular embodiment. Current approaches capitalize on the success of foundation models in the language and visual domains and offline supervised learning from large datasets. To make generalization to different robot bodies possible, their embodiment within the model is drastically reduced and regularized to have similar inputs (images) and similar outputs (position in Cartesian space). Additional modalities, like joint angles both for proprioception (input) and for action (output) can in principle be added (see (Reed et al., 2022)). The visual observations can also be ego-centric rather than coming from external cameras (Yang et al., 2024), which is more in line with embodiment and situatedness (perceiving the environment trough your own sensors and from your perspective). However, every such addition to the model will necessitate a bigger network with limited potential for generalization to other agents controlled by the same model. Therefore, there are intrinsic, fundamental trade-offs in cross-embodiment learn-

ing: the bigger the differences between the robots and the more profound the representation of their sensorimotor space is, the less one can profit from a cross-embodied brain. In fact, I predict that there will be situations where the shared backbone will limit what a particular agent can learn.

Acknowledgments

References

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1-3):139–159.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: MIT Press.

Hoffmann, M. a Pfeifer, R. (2011). The implications of embodiment for behavior and cognition: Animal and robotic case studies. Tschacher, W. a Bergomi, C. (ed.), In *The Implications of Embodiment: Cognition and Communication*, pp. 31–58. Exeter: Imprint Academic.

Lungarella, M. a Sporns, O. (2006). Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2(10):e144.

Müller, V. C. a Hoffmann, M. (2017). What is morphological computation? On how the body contributes to cognition and control. *Artificial life*, 23(1):1–24.

Padalkar, A., Pooley, A., Jain, A., Bewley, A., Herzog, A., Irpan, A., Khazatsky, A., Rai, A., Singh, A., Brohan, A. et al. (2023). Open X-Embodiment: Robotic learning datasets and RT-X models. *arXiv preprint arXiv:2310.08864*.

Pfeifer, R. a Bongard, J. (2006). *How the body shapes the way we think: a new view of intelligence*. MIT press.

Pfeifer, R., Lungarella, M. a Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. *Science*, 318(5853):1088–1093.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J. T. et al. (2022). A generalist agent. *arXiv preprint arXiv:2205.06175*.

Yang, J., Glossop, C., Bhorkar, A., Shah, D., Vuong, Q., Finn, C., Sadigh, D. a Levine, S. (2024). Pushing the limits of cross-embodiment learning for manipulation and navigation. *arXiv preprint arXiv:2402.19432*.

Sketch Map Representation of Cognitive Graph

Petr Rapant[1], Marek Menšík[2], Adam Albert[2]

[1]Department of Geoinformatics, Faculty of Mining and Geology, VSB - Technical University of Ostrava,17. listopadu 2172/15, Ostrava–Poruba, 70800, Czech Republic
[2]Department of Computer Science, Faculty of Electrical Engineering and Computer Science, VSB - Technical University of Ostrava,17. listopadu 2172/15, Ostrava–Poruba, 70800, Czech Republic
petr.rapant@vsb.cz, marek.mensik@vsb.cz, adam.albert@vsb.cz

Abstract

Cognitive sciences deal, among other things, with studying the representation of space in the human brain. In general, they work with two possible models of the internal representation of the environment: a cognitive map and a cognitive graph. The cognitive map is embedded in a global Euclidean framework and allows for determining directions and distances between objects of interest. In contrast, the cognitive graph primarily represents the environment's topology; it does not contain global metric data, but it still allows, for example, to search for paths between objects of interest. To study both representations, researchers use test subjects to walk through an unfamiliar environment (either real or virtual) and then ask them to verbally describe their journey through the test environment or draw a sketch map of the environment. In principle, they thus externalize their internal representation of the environment. The created representations are then analyzed for research purposes. In cognitive research, the first or second form of presentation of acquired knowledge is usually used. Of course, each has its advantages and disadvantages.

Interestingly, however, both forms are not traditionally used simultaneously. Our research team deals with the possibility of obtaining spatial data from text descriptions and their automatic graphic visualization in the form of sketch maps. Since the textual description is usually simplified and does not contain global metric data, the resulting sketch map is more of a cognitive graph created in the subject's brain than a cognitive map.

The developed procedure will also be able to be used interactively, making it possible to interactively correct the text description so that the created sketch map best reflects the representation of the studied scene in the subject's brain. The image obtained for the internal display can thus be more faithful. The procedure we developed could thus enable the connection of both representations (graphical and textual) and, in the case of cognitive research, a better understanding of the internal representation of knowledge about the researched environment. In our paper, we outline the methods by which this procedure can be implemented, which are primarily based on natural language processing and logic methods and show the first results of our research.

At the input of our processing is a narrative – a systematic description of the route of movement through the studied space, which is analyzed by a wide range of natural language processing methods to extract spatial data from the narrative in the form of a well-structured representation. The result represents each sentence of the narrative with a structure [from where, where, where, how]. From sentences represented in this way, it is possible to construct a graphical representation, for example, in the form of a knowledge graph or labeled planar graph. Computer processing tools quickly grasp the latter. The result of labeled planar graph processing is a sketch map.

Investigating Evaluation Techniques
for Dynamic and Interactive Cartographic Stimuli

Tomas Vanicek[1], Marketa Beitlova[1],
Michaela Vojtechovska[1], Stanislav Popelka[1]

[1] Department of Geoinformatics,
Faculty of Science, Palacký University Olomouc, 17. listopadu 50, 771 46 Olomouc, Czechia;
tomas.vanicek@upol.cz, marketa.beitlova@upol.cz, michaela.vojtechovska01@upol.cz, stanislav.popelka@upol.cz

Abstract

The increasing emphasis on improving the usability of digital tools, with a particular focus on interactive maps, underscores the importance of understanding user requirements and the need for continuous improvement of the interactivity of digital interfaces. Despite significant efforts to satisfy a wide range of user preferences and needs, obstacles remain in effectively using these maps, particularly regarding their design, content, and controls. This study investigates the complexities associated with evaluating the user-friendliness of maps that change in real-time instead of static ones, acknowledging that the former presents a more complex challenge. This paper outlines the difficulties associated with using interactive maps versus static images in studies that utilize eye-tracking technology. It presents possibilities for evaluating dynamic maps based on findings from two different eye-tracking experiments. The first study used a stationary eye tracker with a digital interactive atlas as the stimulus. In contrast, mobile eye-tracking glasses and a static paper atlas were employed in the second study. The practical part of this study details how data from both experiments were analyzed to understand user interactions with the maps, focusing on a comprehensive comparison. Regarding data analysis, analyzing static images tends to be straightforward due to the unchanging nature of the content, which allows for easy identification of the participant's gaze points. However, analyzing user interaction with changing maps is somewhat more complicated due to the variable content, posing a significant challenge. The paper thus proposes future research directions to discover new methods for evaluating dynamic maps and interactive elements, which will enhance the usability assessment for dynamic and interactive maps about eye-tracking experiments.

The Expanded Mind as an Educational Challenge: towards new forms of Creating Sociotechnical Systems in the educational Process

Michal Černý

Faculty of Arts, Masaryk University, Brno
Arne Nováka 1, 602 00, Brno, Czech Republic
mcerny@phil.muni.cz

Abstract

Developing generative systems with artificial intelligence using large language models has fundamentally transformed the way of work, the labour market, information retrieval and evaluation, the possibilities of solving selected problems, and many other areas. It turns out that the concepts of competencies for communication and collaboration or problem-solving cannot be limited to human-human interactions but must be extended to building dynamic socio-technical systems that will fold to solve specific challenges or tasks. This contribution will focus on transforming education concerning developing students' competencies to enter into, critically evaluate, use and ethically reflect on these socio-technical systems. The manifestation of this change is the concept of AI literacy, which we understand as a fundamental starting point for the formation of ethical, effective and critically reflective socio-technical systems in an educational context.

1 Introduction

The traditional understanding of education is based on the Greek tradition, emphasising the importance of the educated individual. Education refers to the individual who develops his wisdom as a capacity to participate in the life of the polis. The development of educational ideas worked further with this individualised form; it can be encountered in medieval universities, where the award of an academic degree corresponded to some form of ontological transformation or promotion to nobility, but also in other areas of education (Keller & Tvrdý, 2008). Traditionally, students have knowledge and skills in particular subjects; education focuses on measurable learning outcomes. All these aspects point towards educating individuals (students) and equipping them with other entities (knowledge, skills, attitudes). Lakoff and Johnson (1999, 2003) might say that the education process corresponds to a container metaphor, where the student is the container and the individual educational contents are the objects put into the container.

The development of digital technologies and second-generation cognitive sciences brings the first important aspect of a change of perspective on the constitution of the primary forms of education. According to Siemens (2005), knowledge is not entities but connections in a network. The environment of problem-solving, education or science increasingly emphasises collaboration and building networks that dynamically emerge according to the task (Clark & Chalmers, 1998; Rupert, 2010). Thus, the focus of education is not an entity or an individualised being but a network that can establish dynamic relationships with other elements (inforgs) in the information environment (Floridi, 2014, 2015), depending on what challenges it is currently facing.

The development of this new educational paradigm occurs only in the context of the information revolution, which, according to Floridi, brings another critical aspect, namely, the blurring of the boundaries between technical and human elements that interact informationally (Floridi, 2014, 2019). It seems necessary to speak of education in this sense as a transformation from a purely humanistic phenomenon, emphasising the role of the individual human being, to a socio-technical paradigm (Appelbaum, 1997). In the environment of information interactions, what is often important is not whether the human being or the technology is part of it but the building of a typical interaction base, the formation of new systems that make it possible to solve problems with a significantly higher degree of complexity and intricacy than was previously possible.

The development of artificial intelligence in the last two years (Abdullah et al., 2022) has substantially increased the demands on building these new systems, understanding thinking and problem-solving, not as an individualised process at the level of entities but as a dynamic manifestation of networked interactions. The abilities of individual network elements to connect and exploit interactions with artificial intelligence appear to be an essential characteristic of their competence equipment.

2 AI literacy

Systems using AI are not value-neutral (Bridle, 2018). Heidegger's notion (Cerbone, 1999) of the neutrality of technology (the famous example of the hammer) is dynamically changing, affecting spirituality, ethics, the

labour market and many other areas (Baidoo-Anu & Owusu Ansah, 2023; Kőkuti, 2023; Latzer, 2022; Reed, 2021). The complexity associated with their development leads to the need to shape AI literacy as a particular skill for working with such systems. Ng et al.(2023; 2023) mention four areas they believe this new literacy cuts across. In the list below, we supplement the venom with a brief comment on our notion of educational transformation.

Knowledge of the general principles and principles of how artificial intelligence works
The ability to understand selected limitations of technologies and how they work can be highlighted in this area. We do not believe that detailed principled knowledge is essential, but rather the ability to work with the limits of technologies so that their integration into socio-technical schemes is effective and leads to sufficiently robust and reliable outcomes.

Ability to use selected tools with artificial intelligence
This area involves skills related to working with prompts, refining search queries, asking questions, etc. At the same time, it also includes a general understanding of the available tools and the ability to search for and test them. In this area, we can talk about specific tool literacy, allowing us to develop individual tools' possibilities fully.

Ability to design and create new tools and systems with artificial intelligence
This area can be understood in two ways. Firstly, it can be about actually designing new systems and implementing them, but we consider this more of an engineering skill. In our conception, it is more about the ability to implement existing solutions appropriately in a socio-technical system, to ensure the permeability of results between the different elements of the system, and also to customise individual solutions where commonly available skills (GPTs) are involved.

Ability to ethically evaluate and reflect on AI systems
This area requires a fundamental rethink of some ethical considerations related to individual ethics. This forms only a tiny part of the necessary moral and social analysis. The emphasis on social ethics will represent one of the vital transformational elements that will be increasingly important in contemporary knowledge and for which most students are unprepared.

3   Conclusion

The above considerations show that education faces two daunting tasks: 1) abandoning the individualistic entity model in favour of a network model and 2) developing AI literacy as part of everyday education at all levels in most school subjects or courses. AI literacy does not form a stable corpus of learning outcomes; it is

dynamically variable, as is the entire development of AI in the last two years. This makes it challenging to grasp and implement it into mainstream education. Nevertheless, without it, ensuring responsible and meaningful building of socio-technical systems is impossible, a fundamental prerequisite for employment in today's society.

These four components of AI literacy can be a critical educational starting point. They show that we need to leave behind the idea of education as an individual performance. More valuable is the ability to create dynamically formed networks and hybrid communities that students learn to form as they work on individual problems. Related is the emphasis on ethics as a social phenomenon rather than an expression of individual virtue.

Students will have to learn to master not only the user's work with individual tools but, above all, to acquire the ability to think about problems in a new way, to solve them, to (partially) algorithmise them, or to emphasise not analytically exact but iteratively approximate solutions to problems.

Literature

Abdullah, M., Madain, A., & Jararweh, Y. (2022). ChatGPT: Fundamentals, Applications and Social Impacts. *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 1–8. https://doi.org/10.1109/SNAMS58071.2022.100626 88

Appelbaum, S. H. (1997). Socio-technical systems theory: An intervention strategy for organizational development. *Management Decision*, *35*(6), 452–463. https://doi.org/10.1108/00251749710173823

Baidoo-Anu, D., & Owusu Ansah, K. (2023). *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id= 4337484

Bridle, J. (2018). *New dark age: Technology and the end of the future*. Verso Books.

Cerbone, D. R. (1999). Composition and Constitution: Heidegger's Hammer. *Philosophical Topics*, *27*(2), 309–329. JSTOR. http://www.jstor.org/stable/43154325

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L. (Ed.). (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Springer.

Floridi, L. (2019). *The logic of information: A theory of philosophy as conceptual design*. Oxford University Press.

Keller, J., & Tvrdý, L. (2008). *Vzdělanostní společnost? Chrám, výtah a pojišťovna*. Sociologické nakl. (Slon).

Kőkuti, T. (2023). Artificial Intelligence in a Transforming Labour Market–New Skills are Needed? *Journal of Recycling Economy & Sustainability Policy*, *2*(1).

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought (Nachdr*. Basic Books.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. The university of Chicago press.

Latzer, M. (2022). The Digital Trinity—Controllable Human Evolution—Implicit Everyday Religion: Characteristics of the Socio-Technical Transformation of Digitalization. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *74*(S1), 331–354. https://doi.org/10.1007/s11577-022-00841-8

Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S., & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. *Education and Information Technologies*, *28*(7), 8445–8501. https://doi.org/10.1007/s10639-022-11491-w

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, *71*(1), 137–161. https://doi.org/10.1007/s11423-023-10203-6

Reed, R. (2021). The theology of GPT-2: Religion and artificial intelligence. *Religion Compass*, *15*(11), e12422. https://doi.org/10.1111/rec3.12422

Rupert, R. D. (2010). *Cognitive systems and the extended mind* (1. issued as an Oxford Univ. Press paperback). Oxford Univ. Press.

Siemens, G. (2005). Connectivism: A theory of learning for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*(1).

Exploratory approaches to analyzing eye-tracking data for multiple-choice questions: An example of graph interpretation

Martina Kekule

Charles University, Faculty of Mathematics and Physics
V Holesovickach 2, 180 00 Praha, Czech Republic
martina.kekule@seznam.cz

Abstract

We used the eye-tracking method to observe 32 high school students' strategies when they solved 7 multiple-choice tasks focused on graph slope in kinematics. We provide comparison of attention allocation between different students' groups: those, who solved a test task correctly/incorrectly and those who solved the whole test the best/the worst. Moreover, we applied the cluster analysis in order to divide students into groups according to their attention spent on each option and a task stem. The groups of students so selected differed most in the time they spent reading the stem or the stem graph.

1 Introduction

Eye tracking research has a long history, but it is still a relatively new method in physics education research. A common approach is to focus on solving multiple-choice tasks and provide a comparison between correctly and incorrectly answering students; typically operationalized the expert-novice paradigm in education. It may also be useful to focus on exploratory statistical techniques. Specifically, we illustrate how to use cluster analysis, and look for factors that could possibly indicate differences in students' strategies when solving tests.

Earlier, we focused on differences in gaze allocation between correctly or incorrectly answering students (Kekule, 2019). In this contribution, we compare the extremely performing students (i.e. those who solve the whole test the best) to the non-extremely performing students (i.e. those who solved a task correctly but they did average on the whole test). Secondly, we apply cluster analysis on eye movements data in order to divide students into two groups and look for differences between the two groups.
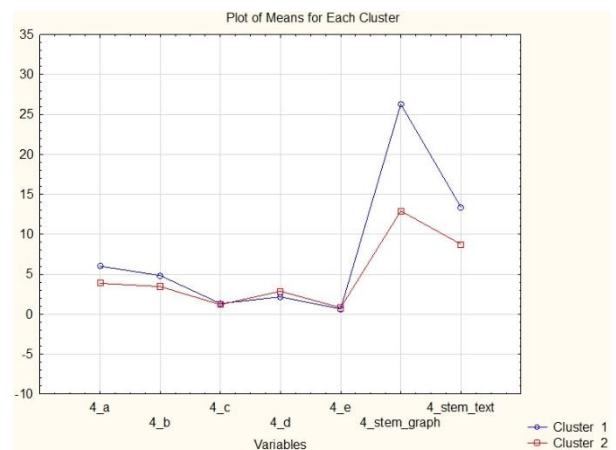
2 Method

32 participants solved 7 multiple-choice tasks focused on the graph slope in kinematics, mostly adopted from TUG-K test by R. Beichner. Eye-tracker TX300 (300 Hz) and Tobii Studio 3.2 by Tobii was used in order to get the fixation data. Stem-text, stem-graph and each option was marked as a separate AOI.

3 Results & Conclusion

We compared total fixations duration mean on an AOI of correctly answering students to the group of the 12 % best performing student. The best performers spent less time on most of the AOIs (options). This was not the case for two -the only quantitative- tasks. We also compared incorrectly answering students to the group of the lowest performing student. The latter group was more focused on options which showed typical "graph as a sketch of a real situation" conception.

We applied the cluster analysis method (k-means approach) to the fixations (count) spent by all students on each AOI and divided the students in two groups. Then we compared means of the fixation count spent on each AOI for every students group (See e.g. Fig. 1).

Fig. 1. Task 4, the average number of fixations of the two groups of students (by k-means method) on each AOI.

For all tasks, the biggest difference between the two groups was in attention to the task stem or stem graph

(See e.g. Fig. 2). There were 25% and 29% of correctly answering students in the groups, so the groups were similar in terms of correct answers. These exploratory results suggest that there might be a difference in students' expertise that would be reflected in the interpretation of the stem task. Moreover, we applied and will discuss the cluster analysis only on data on tasks options.

Variable	Analysis of Variance (data-cluster)	
	F statistics	signif. p - value
4_a	4,32739	0,047123
4_b	1,96027	0,172871
4_c	0,11174	0,740757
4_d	0,20289	0,655994
4_e	0,41578	0,524492
4_stem_graph	24,94574	0,000031
4_stem_text	10,70768	0,002918

Fig. 2. Task 4, results of the inversed ANOVA, the biggest F difference is for the stem and the stem graph.

Literature

Ding&Beichner R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics – Physics Education Research* 5.

Kekule, M. (2019). Students interpretation of graph slope in kinematics investigated by the eye-tracking method In Levrini&Tasquier (Eds.): *Electronic Proceedings of the ESERA 2019 Conference,* pp. 34-3, Bologna.

User testing of information presentation using eye-tracking

Gabriela Godišková

Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00, Brno, Czech Republic
gabriela.godiskova@phil.muni.cz

Abstract

This study examines user experience across three online data journalism articles using eye-tracking and retrospective think-aloud. Findings reveal that participants tend to follow similar transition paths within the linearly structured article, however in non-linear articles there were various transition strategies revealed. Participants also revisited data visualisations and their annotation texts to clarify information. Notably, animated data visualisations included in two of the articles posed challenges; they caused confusion, delays and sometimes were mistaken for advertisements.

1 Introduction

User research in the field of data journalism has not been explored to a great extent so far (Errey, et al., 2023; Riche et al., 2018). Multiple factors have contributed to this, such as time pressure or a lack of funding for data journalists. The presented research studies user transitions through selected data journalism articles. Emphasis was placed on exploring the perception of relationships between data visualisations and their text annotations, as well as the effect of specific data visualisations on overall user experience with data journalism articles.

2 Methodology

The eye-tracking method was selected to perform the study, complemented by a retrospective think-aloud protocol. The eye-tracking research was performed using the device Tobii Spectrum 300. Three data journalism articles by the Czech online medium iRozhlas were used as stimuli. Ten participants aged 30-39 participated in the research, and they were instructed to freely interact with articles as they would in normal life.

Selection of the articles was based on their structure – primarily, they differed in the order of the contained elements. Two of the three articles were structured in linear order, which means that all the article elements such as title, lead paragraph, main text, data visualisations or images were placed in an uniform order

below each other. The remaining third article was structured in a non-linear way meaning that the article elements mentioned above were irregularly arranged – for example, next to the main text there were boxes with additional text.

Additionally, the articles differed in the data visualisation dynamics. They contained either static or animated data visualisations. The animation created within the data visualisation was responsible for its change in appearance over time – for example a scatter plot that changed visualised data within different years. The animation was presented in an automatic loop with no user control of playback.

User outputs were subsequently analysed by dividing articles into areas of interest (AOI) regarding the article elements.
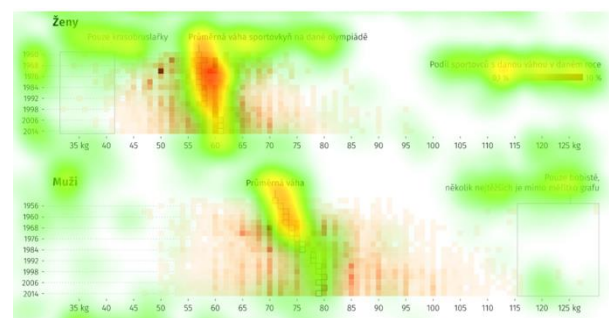
Fig. 1. Participants' fixations following the annotation elements within data visualisation from the first article.

3 Main findings

Participants interacted with the first and second presented stimuli in a mostly linear manner. In each article, eye-tracking data exhibited frequent returns between data visualisations and their annotation texts or other relevant surrounding text, with participants citing misunderstanding or the desire to reinforce presented information as the reason. Annotation texts were generally evaluated positively by participants, helping to diffuse confusion around the meaning of visualisations (see Fig. 1). The third article displayed a larger variance in transition paths between participants, due to containing an extra text box that some participants read

and some chose to ignore, stating that the box appeared as advertisement.

Animated data visualisations were viewed as rather problematic by participants. In the first article, the speed of the animated visualisation caused participants to spend more time waiting for the playback to restart. In the third article, the animated visualisation was even mistaken for an advertisement and completely ignored by some participants.

Consistency in data visualisation types was evaluated differently based on context. The first article combined three similarly looking visualisations with several others, which tended to disorientate participants. However, the identical style of visualisations throughout the whole second article was described as easy to follow and understand.

Additionally, participants reported an interest in an exploration of a correlation diagram in the second article, as well as a map visualisation in the third article, expressing an expectation of interactivity based on their previous experience with similar visualisations online.

4 Discussion

The research showed that author-driven linear transition through articles was mostly followed by participants when present, i.e. in the first and second article. This is in accordance with Kartveit (2018). Explanative annotations were evaluated as helpful by participants, confirming the findings of Thudt et al. (2018) and Franconeri et al. (2021).

Findings diverging from previous research concern the participants' evaluation of animated visualisations; while Chevalier et al. (2016) found these to be improving the overall user experience, participants of this study reported them as problematic and confusing.

Further research regarding the perception of animated visualisations, as well as the user transitions in non-linear article structures such as the third stimulus studied here, can be a reasonable next step to perform using quantitative methods with a larger number of participants.

Acknowledgements

Literature

Chevalier, F., Riche, N. H., Plaisant, C., Chalbi, A., & Hurter, C. (2016). Animations 25 Years Later. *Proceedings Of The International Working Conference On Advanced Visual Interfaces*, 280-287. https://doi.org/10.1145/2909132.2909255

Errey, N., Liang, J., Leong, T. W., & Zowghi, D. (2023). „Evaluating narrative visualization: a survey of practitioners". In: *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-023-00394-9

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). „The Science of Visual Data Communication: What Works". In: *Psychological Science In The Public Interest* 22.3, s. 110–161. https://doi.org/10.1177/15291006211051956

Kartveit, K. (2018). „They never made it to the end: Reader uses of a multimedia narrative". In: *Journal of Applied Journalism & Media Studies* 7.2, pp. 289–309. https://doi.org/10.1386/ajms.7.2.289_1

Riche, N., Hurter, C., Diakopoulos, N., & Carpendale, S., ed. (2018). *Data-Driven Storytelling*. A K Peters/CRC Press. https://doi.org/10.1201/9781315281575

Thudt, A., Walny, J., Gswchwandtner, T., Dykes, J., & Stasko, J. (2018). „Exploration and Explanation in Data-Driven Storytelling". In: *Datadriven storytelling*. Ed. N. H. Riche. Taylor & Francis / CRC Press.

Navigating through the jungle of international collaborative grants: Lessons learned from ASP-belong project and beyond

Radka Martincová

Faculty of Arts, Masaryk University
Arne Nováka 1, 602 00 Brno, Czech Republic
martincova@phil.muni.cz

Abstract

This contribution aims to provide guidance for researchers who wish to enrich their project portfolio by participating in international collaborative projects. It outlines strategies for achievement of higher success rate in grant competitions and presents several suitable grant schemes, most of them funded by the European Union.

1   Introduction

The most natural way to join international collaborative projects is upon invitation from colleagues, who are preparing a project proposal. This article aims to highlight several other gateways for researchers who would like to strengthen their presence in international consortia, notably those formed around EU funded projects.

2   COST actions

COST actions are large international networks assembled around a particular research topic. The programme funds networking costs only, but many countries (including Czechia) run complementary grant schemes funding research conducted in the defined area. COST actions focus on development of innovative research topics from any field of research. The actions are very flexible: researchers interested in the topic may enter also ongoing actions. Benefits for participants involve broadened network of contacts including potential partners for future project cooperations; many COST actions are translated in successful Horizon bids in future.

2.1   COST: Strategic tip

Browse accepted actions. You may consider joining the action of your interest by contacting the action lead. Example: CA21131 – *Enabling multilingual eye-tracking data collection for human and machine language processing research*.

3   Horizon Europe (2021–2027)

Horizon Europe is the EU's key funding programme for research and innovation supporting individual researchers, research teams and most prominently international consortia.
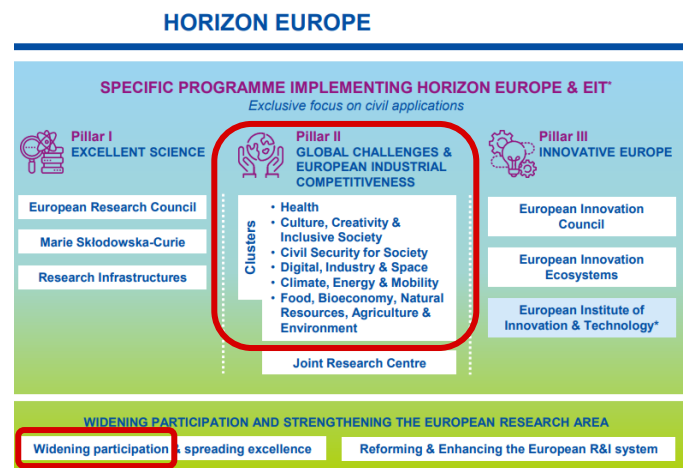
Fig. 1. The structure of Horizon Europe

3.1   "Clusters"

These are the classical EU-funded consortium projects that most researchers have heard of or took active part in. In Horizon Europe, they are assembled into six thematic clusters (see *Fig. 1*). The project calls and topics are listed in Work programmes that are typically issued every two years.

3.1.1   "Clusters": Strategic tips

1) Work programmes for years 2026–2027 have not been published yet, but as preparation of complex proposals typically takes many months, it is important to get ready for application as soon as possible. At time of the official publishing of work programmes, it may be often too late to form competitive consortia. The leaked drafts of the work programmes containing foreseen calls

traditionally appear on-line long before their official publishing.

2) Many newcomers are tempted to take on the role of project coordinators, which may result in frustration as enormous efforts are often wasted. It is advisable to start by joining promising consortia involving experienced partners. The partner search often doesn't yield expected results: it is better to get in via research acquaintances or to approach directly potential leaders in the area. (Analysis of projects funded in past may reveal who the leaders are. When you contact them, you'll more likely succeed if you offer concrete expertise that is demanded by the call).

3) Look at topics *across* clusters. Browse more than one work programme, your expertise may fit in topics that may look distant at the first sight. For example, even very technical topics often demand expertise in social sciences to leverage the societal impact of breakthrough innovations.

4) The basic research does not suffice, applied results and extensive dissemination efforts are expected. The structure of the consortia is rarely (if ever) purely academic.

5) The calls/topics usually contain very specific information about desired contents of projects. Consortia are supposed to address these expectations very carefully: the modification of project's focus according to researchers' interests or expertise often results in low evaluation score.

6) Assuming a minor role in a project can lead to more substantial cooperation in future. Example: A researcher from our department at Masaryk University was invited to join a project proposal ASP-belong on adolescent mental health from cluster Health with a minor set of tasks. Following resignation of a Serbian partner, she was assigned a much more important role with corresponding budget. Due to Brexit, the British leader was unable to act a formal coordinator and Masaryk University became administrative coordinator of the whole consortium.

3.2 Widening participation and spreading excellence

The work programme "Widening participation and spreading excellence" offers number of grant schemes that are intended to close the research gap between research intensive countries and the so-called "Widening countries" whose performance in both research and ability to reach for the EU research funding is lower. (This group also includes Czechia and Slovakia). The thematic focus of these calls is typically not pre-defined: any field of interest is eligible for funding. Originally, these schemes focused mainly on networking and capacity building, but they were recently complemented by purely research activities.

3.2.1 Twinning

Among the Widening schemes Twinning is probably the most popular. A "Widening" coordinator with two or more excellent partners work together to improve the research performance of the leader. The budget up to 1.5 million EUR is intended mainly for networking purposes (workshops, exchanges, summer schools, conferences etc). However, up to 30% of the budget can be used for research. This can be a great experience for teams considering a role of coordinator.

3.2.2 Widening: Strategic tips

1) Observe evolution of success rate and amount of funding allocated to topics that you consider for application.

2) The successful projects are listed at the bottom of pages of closed calls: browsing the projects funded under a call of your interest may provide valuable insights. (Example: Twinning 2023).

3) If you work at a "Widening" institution, check if it already implements such projects. Find out who's the one smart person that can support your application at your faculty, institute or central level. If there is no such person, find your National Contact Point and establish a friendly working relation with them.

4 Seek for hidden opportunities

Even though Horizon Europe is the most prominent EU programme supporting research and innovation, there is number of other opportunities that may result in fruitful collaborations. The Erasmus+ programme offers much more than student exchanges: the capacity building projects can get very close to research cooperation, e.g. while focusing on PhD training. Researchers interested in cooperation with armed forces may profit from participation in calls of the European Defence Fund. Sometimes it's worth to take a risk and apply in brand new programmes and calls: the success rate can be higher than in case of established well-known schemes.

5 Conclusion

You don't have to be very experienced or the best in your field to succeed in international grant competitions. The key is making strategic choices regarding projects, which you decide to join, knowing where to seek help and giving your proposal appropriate attention. The list of grant schemes presented in this article is by no means exhaustive, rather, it shows several different ways to enter the arena of international collaborative grants.

Acknowledgements

Where to start

CORDIS: Research engine for previously funded EU projects.

COST association: official website.

EU Funding & Tenders Portal: Here you can browse all the EU calls for funding (ongoing, past, and foreseen).

Horizon Europe: official website.

National contact points: Horizon Europe support in your country.

Lexical strategies for encoding Chinese dish names: Cross-linguistic stability and variation

Wei-lun Lu

Faculty of Arts, Masaryk University
Arne Nováka 1, 60200 Brno, Czech Republic
wllu@phil.muni.cz

Abstract

In this study, I address the lexical strategies for encoding Chinese dish names in Chinese and a typologically distant language (which is Czech) from a cognitive semantic point of view, with emphasis on the use of conceptual mechanisms such as metaphor, metonymy and image-schemas. The study will be methodologically based on the use of multiple parallel texts, comparing Chinese with a distant language using different versions of menus from sampled Chinese restaurants in Czechia.

The main findings to be presented include: firstly, some metonymy-based expressions in Chinese (such as 雙冬 *shuāng-dōng* and 羅漢齋 *luóhàn zhāi*) simply cannot get through, or barely gets through, to the Czech language; secondly, some homophony-based Chinese dish names (such as 菩提肉 *pútí ròu*) do not get through to the Czech language at all; thirdly, some figurative Chinese dish names are highly obscure and when they do get across to the Czech menus, they keep their highly abstract and figurative nature (such as 三鮮 *sān-xiān*, 五香 *wǔ-xiāng,* 八寶 *bā-bǎo*, and 全家福 *quán jiā fú*); fourthly, in rare cases, Chinese-specific creativity may get through to the target language (or by accident "leak" to the target language, so to speak) from the source language (such as the case of 什錦魚 *shíjǐn yú*), resulting in a highly poetic (though completely inaccurate) construal of the dish.

With the above findings, I discuss the language-specificity of the Chinese linguistic tool against the framework of Radical Construction Grammar and the methodological advantage of studying renditions of Chinese dish names in various random languages following a Multi-ParT approach, in the backdrop of the socio-cultural turn of Cognitive Linguistics.

Literature

Kövecses, Z. 2010. *Metaphor: A practical introduction* (2nd ed.). Oxford University Press.

Kövecses, Z. 2015. *Where metaphors come from: Reconsidering context in metaphor*. Oxford University Press.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Lakoff, G. 1994. What is a conceptual system? *The nature and ontogenesis of meaning*, ed. W. F. Overton and D. S. Palermo, 41–90. Hillsdale, Lawrence Erlbaum.

Lu, W. (2020). Cultural "signs of life" in politics: A case study of eulogistic idioms for Taiwanese politicians. *Signs of life: Changes and continuity in language, thought and identity*, ed. by Vera da Silva Sinha, Ana Moreno-Núñez and Zhen Tian, 139-154. John Benjamins.

Lu, W. (2017). Religion and worldview in metaphor use: Cultural conceptualisations of death in Taiwanese Buddhist and Christian eulogistic idioms. *Advances in Cultural Linguistics*, ed. Farzad Sharifian, 49-64. Springer.

Lu, W. and S. Shurma. 2021. Rituals. *Analysing religious discourse*, ed. Stephen Pihlaja, 217-234. Cambridge University Press.

Sharifian, F. (2011). *Cultural conceptualisations and language: Theoretical framework and applications*. John Benjamins.

Exploring Subjective Religious Imagery: Artificial Intelligence and Eye-tracking Integration

Markéta Muczková, Michaela Vojtěchovská, Stanislav Popelka, Tomáš Bubík, Klára Kubálková, Nikola Svobodníková, Kamila Fačevicová

Palacký University Olomouc
Třída svobody 26, 77900, Olomouc, Czech Republic
marketa.muczkova01@upol.cz

Abstract

This research focused on the testing of an innovative methodology for exploring religious concepts through artificial intelligence (AI) and eye tracking technology. The "AI Visionaries" project examined the visual associations of participants generated using AI. These images were classified according to religious traditions and available data from religious studies. Participants underwent either an online test or a laboratory version with an eye-tracker, testing a new method aimed at potentially supplementing research tools for investigating subjective human concepts.

The aim was to build upon religious theories and validate them in participants' subjective religious concepts. There are various theories, such as people envisioning religious concepts anthropomorphically or holistically, which have not been directly experimentally tested. Some researchers suggest that people tend to imagine the supernatural as anthropomorphic entities, while others argue for a more holistic perception etc.

The project involved generating images for 22 predefined concepts inspired by religious theories and traditions. The image generation process primarily relied on Midjourney, complemented by the incorporation of Microsoft Copilot and GPT4. To ensure the images' alignment with their intended categories (a total of 22), we influenced the pictures' outcomes by specifying the prompts with all the necessary keywords. Although most of the images were generated without significant issues, a subset required manual retouching or alternative generative AI methods due to the challenges encountered. Subsequently, the finalized images underwent categorization according to their corresponding conceptual frameworks.

These concepts reflected various aspects of religion, including deities, sacred spaces, and moral values. Participants were tasked with selecting images that best matched their subjective concepts, allowing us to test the validity of religious theories in practice. Each image contained factors that could influence perceptions of the supernatural, as outlined in the image prompts. Subsequently, each image was assigned factors that experts in religious studies identified as the most influential and defining aspects of beliefs. The total number of factors was 14, and each of the individual images contained all of these factors. These factors were dichotomous, such as East, West, altruism, egoism, etc. On a scale of 0-6, each factor in each image was then evaluated to attribute the obtained score to participants during image selection.

The research focused not only on the form of participants' religious concepts but also on the factors influencing their specific beliefs. Eye movements from the eye-tracker were analyzed to better understand the visualization process and participants' resulting image choices. The order of images, including the order of the concepts themselves, was randomized to prevent bias.

For research purposes, new software was designed to capture real-time data, allowing analysis of eye movements, AOI recordings, and continuous scoring of factors during the experiment. The software evaluated the data while they completed the experimental task. After completing the task, the participants were shown a graph that illustrates the factors that influenced their subjective religious visualizations. This graph displayed which elements of religious concepts were the most preferred and which influenced and shaped the visualizations the most. This approach allowed us to gather data not only on the form of individuals' subjective religious concepts but also on the processes and factors shaping these perceptions.

Acknowledgements

Composing a geometric abstraction with a generative algorithm

Aleš Svoboda

Faculty of Humanities, Charles University
Pátková 2137/5, Praha 8 – Libeň, 182 00, Czech Republic
ales.svoboda@fhs.cuni.cz

Abstract

Formalist aesthetics identified the empirical impairment of pictorial composition as a relevant part of artistic creation. Composition then means the optimal assembly of partial building elements with regard to the apperceptive expectations of the perceiver. The promotion of non-representative art in the second decade of the 20th century and its further development followed primarily this intuitive search for building elements and their most aesthetically satisfying connections. Assuming a finite number of certain elements for a given syntax, rules for their mutual connection can be defined. This can then become the starting point for a mathematically accessible structure. A generative algorithm could provide a method for simultaneously satisfying multiple conditions of the observed "fitness" functions.

1 Non-representative art and its formalist reflection

Non-representative art, which appeared in the art space in the second decade of the 20th century, initially gained its appreciation mainly thanks to the intuitive finding of previously unknown visual attractions. However, the compositional considerations of the abstract artists of the first half of the 20th century, although they followed rather intuition, over time began to call for a rational and deliberate procedure. The idea of possibly finding more impressive works and therefore successful solutions led continuously to the elaboration of many specific rules, as, for example, the pedagogues of the influential Bauhaus modernist school tried to do. At the same time, formalist aesthetics outlined the empirical impairment of pictorial composition as a relevant part of artistic creation. Composition was then understood as the optimal assembly of partial architectural elements with regard to the apperceptive expectations of the perceiver. In this context, it is possible to recall the protagonists of the effort to rationally affect color effects – Johannes Itten (Birren, 1970) and Josef Albers (Albers, 1971). Such an opinion gradually took a quite unquestionable position in the 20th century. The example for all can be, in a sense, the heir to the Bauhaus, August Herbin (1882–1960). He wrote his own method of working with colors, surprisingly based on Goethe's color theory, and

published it in 1949 in the book *L'Art non-figuratif non-objectif*. In any case, such a conceptual approach to geometric composition and its color filling can become mathematically difficult.

2 Computer art and the rational development of the formalist concept

In computers, as universal machines, from the mid-60s at the latest, artists and programmers with creative inclinations began to see reinforcement for visual imagination. Computers, however, contained ambivalence in their character – the deterministic nature could be a support for searching for and fulfilling compositional rules, but at the same time, its rigidity could also seem counterproductive to spontaneity as an assumed necessary part of creativity. However, this strictness soon seemed to
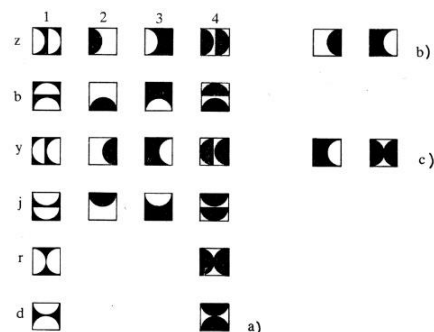
be overcome by the inclusion of randomness.

Obr. 1. Fig. 1. Sýkora's initial elements of the black-and-white structure from 1966, arranged according to optical density (a) and the possibility of mutual connection either in color or in shape (b, c).

In his pioneering work in the mid-1960s, the Czech painter Zdeněk Sýkora tried to use the computational consistency of a computer program to fulfill clear compositional rules, which, however, were not trivially unambiguous (Sýkora and Blažek, 1970). The set elements of the composition evaluated according to the shape and at the same time the optical density (Fig. 1) were to be applied in the intended composition with regard to the specified neighborhood principles. The sought-after neighboring element could have either a

higher (+) or lower (-) optical density and could be connected to the interface of the elements either with the same color or the same shape. The starting point for a specific calculation was a sketch with the location of several initial elements and a proposal for changes in saturation (Fig. 2 on the left). The algorithm worked out a solution according to which the final image was painted (Fig. 2 on the right). The artistic inspiration here was the rational Cézanne method of painting. However, the application of the rules for element relationships was based only on their linear chaining in the area. Even so, one could speak of an independent model of behavior,
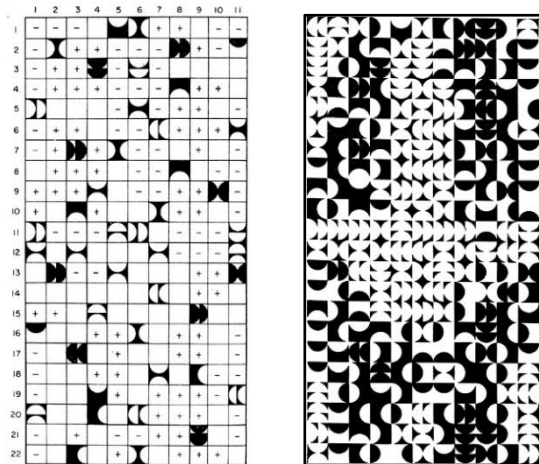
progressing towards autonomous decision-making and intelligence.

Fig. 2 Positioning of the initial elements and determination of the required change in saturation for the inserted adjacent elements for Black and White Structure from 1966 (left) and Zdeněk Sýkora: Black and White Structure, 1966, oil on canvas, 220 × 110 cm (right)

3 Generative algorithms as inspiration

In the 1950s and 1960s, a number of computer scientists independently investigated evolutionary systems with the intention that evolution might be applicable as an optimization tool in engineering problems (Mitchell 1998: 2). John Holland is considered to be the discoverer of genetic algorithms in the 1960s, who, together with colleagues and students at the University of Michigan, developed them further in the 1960s and 1970s. Its original goal was to study adaptation, not to design algorithms to solve specific problems. However, it is precisely this ability to develop programs that, although subjected to random mutations, are subsequently brought to meet the intended goals thanks to the continuous evaluation of their usefulness (fitness functions). In principle, this involves the exchange of parts of "parent" programs and the gradual creation of generations of successors. The modification of this idea, namely the random generation of the internal arrangement of the composition limited by the fitness function, which rejects non-productive variants and expands the advantageous ones, led me to work on a

program (in the Processing language) that would solve the composition with specified rules.

4 Series System 25XT3

Since 2012, I have been incorporating the golden ratio into my generative computer art series. In 2016, it was the Te-XT rumor and Te-XT rumor automat series, which consisted of producing specific square "signs" and then connecting them in an area. The signs were created by layering black and white circles and squares inscribed in a grid, which divided the initial square surface with three axes exactly in the proportions of the golden ratio. The System 25XT3 series then began to include gray as well, and allowed the connection of adjacent elements only while respecting the internal division in the emerging neighborhood in both directions. In other words, the given computer program (System 25XT3, 2017) randomly chose an element respecting the stated rules for the center of the composition and added other elements to it from the center in corresponding proportional divisions. The composition thus grew with regard to the previously
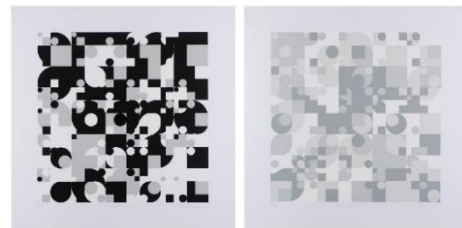
randomly selected axes of work. In this series, in 2017, I realized black and white screen prints System 25XT3 17-03-18 and System 25XT3 17-04-27 (see Fig. 4).

Fig. 4. System 25XT3 17-03-18, 2017, screen printing, paper, 700 × 700 mm and System 25XT3 17-04-27, 2017, screen printing, paper, 700 × 700 mm.

Because I wanted to preserve the possibility of transformation in the final artefact, compositions of separate elements on a magnetic background became a variant in this series (Fig. 5). There was also a color variant. In magnetic reliefs, however, it is not possible to guarantee assembly with regard to the interconnection of neighborhoods.

Fig. 5. Variable magnetic collages: System 25XT3 BW 17-07-10, 2017, variable collage, digital print, paper, magnetic elements, 70 × 70 cm and System 25XT3 COL 17-07-20, 2017, variable collage, digital print, paper, magnetic elements, 70 × 70 cm.

The whole composition can also be imagined as a connection of three layers, where the basic layer consists of 5 × 5 fields, which can be square or circular. In the higher layer, each field is divided in the rate of golden ratio, the random selection of the order of the large and small section is the same for each row and column. In the highest layer, the same division is carried out in the resulting sections, again randomly and identically for entire rows and entire columns. The circles and squares
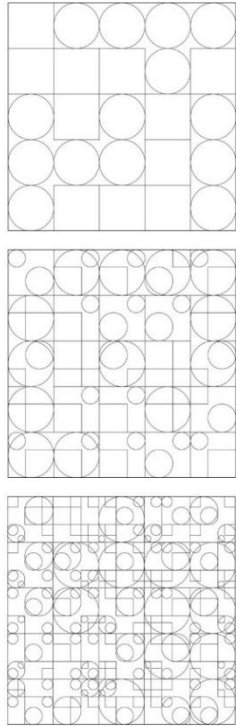

inscribed in the network of golden ratio always overlap the circles and squares of the lower layers (Fig. 6).

Fig. 6. Scheme of three layers in the program for the System 25XT3 series. Basic with 25 fields, higher with division of fields by the golden ratio and the highest layer with further division of the resulting rates again by the golden ratio.

In the System 25XT3 series, the shape arrangement is governed by the rule, but the choice of colors and gray scale were obtained randomly. In the case of using a scale of 12 colors, it is possible to calculate the number of random color combinations of the base surface as 12^{25}, the second layer 12^{50} and the third layer 12^{150}. Their combination could then be 12^{187500}.). However, more important from an artistic point of view is the possibility of producing species-specific visual experiences, which results from the mechanisms of the psychology of perception and the practical realization of a limited type of distinguishable color contexts.

# 5 System 25XT3 based on a program inspired by genetic algorithms

A common modernist consideration of the aesthetic preference for contrasting color relationships and color neighborhood relationships can be used to guide the random selection of colors in the composition created by
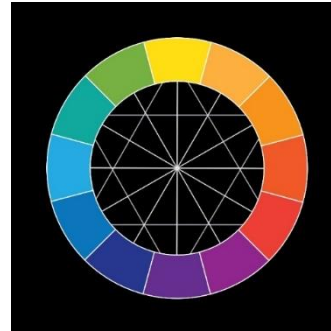
the System 25XT3 rules. Itten's color wheel uses a scale of 12 colors (Fig. 7), i.e. the three basic, primary colors, blue, red, and yellow are complemented by the secondary colors, violet, orange, and green, between which there are another six intermediate degrees.

Fig. 7. Color wheel with 12 tones according to J. Itten.

The traditional lesson claims that opposite, contrasting colors in a wheel are not only more impressive, but also more pleasant, as well as adjacent colors do not cause displeasure to the observer, therefore it is assigned a "0" in the program for their rating, "1" in other cases (Fig. 8).
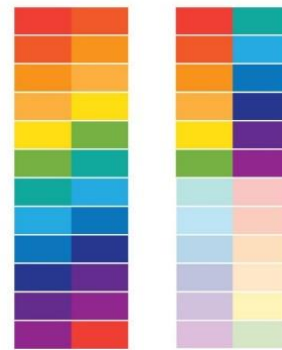
Fig. 8. Color relationship evaluation table according to Itten's wheel: preferred relationships are adjacent colors (12 options on the left) and completely opposite, contrasting colors (6 options on the right).

In the base layer, each field is confronted with its surroundings, and a maximum of 8 relationships (H1, H2, H3, H4, H5, H6, H7, H8) can take on a total value of H = 0 (suitable color) to H = 8 (all relationships do not comply). The algorithm randomly chooses one of the 25 fields, which, depending on the "fitness" result, will either keep its color or change it. Similarly, the color relationship of individual shapes in higher layers to their background in lower layers is evaluated. It is, of course, a question of what can be called one "generation" when the

partial components of the composition are generated sequentially, however, repeated evaluation of the fitness functions verifies the shape of the composition, when after about 200 program loops a complete exchange of
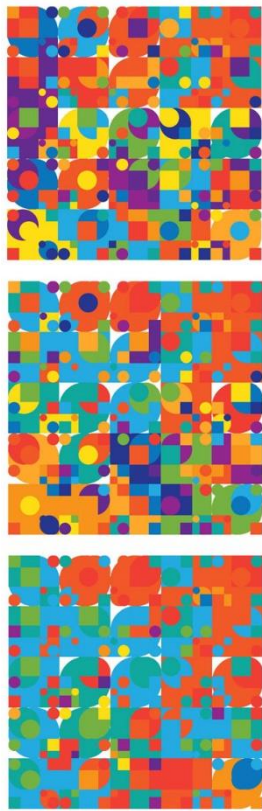
parts can occur (Fig. 9).

Fig. 9. Flow of one process of generating the resulting composition from a completely random assembly through generation "75" to the result in generation "141".
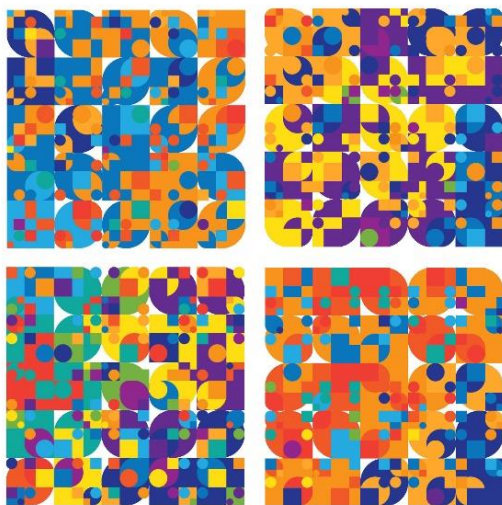
When the process is repeated, it can be seen that from a visually very similar beginning, which seems indistinguishable at a normal glance, it is possible to reach radically different results (Fig. 10).

Fig. 10. Four different results of an algorithm that works by gradually improving the original completely random color neighborhood based on its continuous evaluation.

6 Conclusion

The System 25XT3-GA compiled program works with an initial fixed shape selection and placement setting. The spread of the 12 colors is completely random at the beginning. Gradually, the composition is subjected to an evaluation of the suitability of color neighborhoods, when neighborhoods evaluated positively tend to remain, in neighborhoods evaluated negatively, the color relationship gradually continues to change. The color of the basic 25 fields is always evaluated with regard to its surroundings, i.e. 8 neighboring fields, the color of the shapes that form the neighborhood of the shape in the lower layer is evaluated from the point of view of this relationship. Too loose determination of the rules in the basic layer of course brings the solution relatively quickly, but the result is not particularly specific (the impression of chaos remains); too hard set rules extend the number of generation indefinitely. In some cases, part of the composition may end up in a periodic loop.

The still open topic of computer generative art remains when such a program can be called a manifestation of artificial intelligence. An expert system that makes decisions in the "search space" - that is, actually chooses the right answers - is qualified as intelligent, but in the creative domain, intelligence is only associated with the search for possible new question formulations.

Acknowledgements

Literature

Albers, J. (1971). *Interaction of Color*. New Haven & London: Yale University Press, 2nd ed.

Birren, F., (Ed.)(1970). *Itten, The Elements of Color, A Treatise on the Color System of Johannes Itten based on his book The Art of Color*. New York: Van Nostrand Rainhold Company, 1st ed.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge & London: A Bradford Book, 1st ed.

Poli, R., Langdon, W. B. & McPhee, N. F. (2008). *A Field Guide to Genetic Programming*. Stanford Universiy, 1st ed.

Sýkora Z. & Blažek J. (1970). Computer-Aided Multi-
    Element Geometrical Abstract Paintings, *Leonardo*, 3
    (4): 409-413

Evaluating the Effectiveness of Brand Placement in Sticky Header Element

Stanislav Mokrý, Michal Pšurný, Bilguun Bayarmagnai, Miloš Uldrich

Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno,
Zemědělská 1, 613 00 Brno, Czech Republic
xmokry@mendelu.cz, michal.psurny@mendelu.cz, xbayarma@mendelu.cz, milos.uldrich@firma.seznam.cz

Abstract

Our study evaluated the impact of sticky web elements (fixed elements such as navigation bars and brand logos) on brand awareness and user engagement using a native ad article in two versions: with and without a sticky element at the top. Involving 100 participants, we faced challenges in tracking engagement due to different scrolling behaviors. Using the automatic scroll compensation feature of the EyeLink Data Viewer, developed for WebLink software experiments, we were able to accurately measure attention and interaction, overcoming variability in scrolling positions.

1 Introduction

Sticky web elements, or fixed-position elements, are an integral part of modern web design, enhancing the user experience by ensuring persistent visibility and accessibility. These elements, which include navigation bars, brand logos and call-to-action buttons, remain fixed on the screen as users scroll through a web page (Sikos, 2014). The constant presence of these elements should potentially increase brand awareness. However, Carver (2014) cautions that from a UX perspective, the use of fixed elements should be approached with caution due to their rigid interaction with the user environment. Our eye-tracking experiment aimed to quantify the attention brands receive in such locations and to assess subsequent user reactions.

2 Material and methods

The experiment was conducted at the Mendel University in Brno in the ETLab[1] using the Eye-Link 1000 Plus eye-tracking device, with the sampling frequency set to 500 Hz. The stimulus used was a web article containing native advertising. A total of 100 participants (50 women) were recruited and divided into two equal groups. One group (A) was presented with the stimulus without a sticky header featuring the brand of the native article's advertiser, while the other group (B) viewed the stimulus with a sticky header application.

Participants' eye movements were recorded and analysed to assess their attention to the brand elements within the native ad article. The experiment was facilitated using SR-Research WebLink software, and data analysis was performed using SR-Research DataViewer and RStudio.

3 Results

We utilized the Scroll Compensation feature of SR-Research DataViewer to evaluate measurements differently for stimuli with and without the sticky header, aiming for accurate gaze tracking. The sticky header's presence or absence significantly affected participant attention to the brand in the webpage header, as indicated by fixation counts (11 in version A without sticky header vs. 21 in version B with it) and a Chi-square test p-value of 0.023. However, differences in Dwell Time between groups A (169.88 ms) and B (335.59 ms) were not statistically significant (t-test p-value of 0.057). Post-experiment questionnaires showed no significant difference in brand recall between groups (23 in A vs. 27 in B, Chi-square p-value of 0.4236).

4 Conclusion

In conclusion, the use of Sticky Header elements appears to positively influence initial user attention towards the brand. However, the effects on sustained engagement and brand recall are less clear, highlighting the need for a multifaceted approach in web design and advertising strategies to effectively build brand awareness and engagement. Future research should explore additional factors that may influence the effectiveness of Sticky Headers and other web design elements in advertising contexts.

Acknowledgements

---

[1] https://etlab.cz/

Literature

Carver, M. (2014). *The Responsive Web*. Simon and Schuster.

Sikos, L. (2014). *Web Standards: Mastering HTML5, CSS3, and XML*. Apress.

Simultaneous Eye-tracking and Event-Related Potentials (ERP) Measurement for Consumer Perception Studies

Michal Pšurný, Stanislav Mokrý, Bilguun Bayarmagnai, Jana Stávková

Department of Marketing and Trade, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
michal.psurny@mendelu.cz, stanislav.mokry@mendelu.cz, jana.stavkova@mendelu.cz.
bilguun.bayarmagnai@mendelu.cz

Abstract

The paper aims to show how connecting the eye-tracking method with the electroencephalography (EEG) method is possible—specifically, measuring event-related potentials (ERP) during the simultaneous eye-tracking measurement for consumer studies. Simultaneous measurement is important both from the point of view of the human-brain interface and from the point of view of understanding psychological processes in the brain. Data from both modalities were measured on a sample of 9 participants. The paper will show hardware and software solutions. Furthermore, the preprocessing data and procedure by which ERP data is evaluated and how to connect it with eye-tracking data.

1 Introduction

Consumer neuroscience (neuromarketing) is a field that is gaining more and more attention from marketers. It seeks to introduce new modern methods. Electroencephalography is a neuroscience method that seems to be the best for use in marketing because it is relatively cheap, and it can be deployed even in real conditions such as a store. One way to measure and evaluate this data is event-related potentials (ERP) measurement. The method is commonly used in medical and neuroscience research. The authors try to use this method in consumer perception research (Lin et al., 2018a).

Consumption can be divided into utilitarian and hedonic (Hu et al., 2020). Utilitarian consumption is more functional, cognitively driven, goal-oriented, and practical. In contrast, hedonic products are focused on experience and excitement, are more emotion-driven, and are associated with pleasurable attributes (Herbas et al., 2011).

The objective of the study is to find out if there is a change in selected ERP components when hedonic or utilitarian products are presented with selected marketing tools (discounts and reviews).

ERP components P200 and LPP (Late Positive Potential) are evaluated. P200 is related to early automatic attention, possibly working memory (Donoghue and Voytek, 2022; Lin et al., 2018b). The LPP component is related to affective perception and can serve as an index of later emotional valence, i.e. indicative of the degree of positivity or negativity (Mastria et al., 2017; Pozharliev et al., 2015).

2 Methodology

The data were measured at the Mendel University in Brno in the ETLab[1]. 9 participants were measured. SR Research EyeLink® 1000 Plus eye-tracker was used. LiveAmp 32 amplifier (Brain Products) was used together with an EEG cap 32 channel configuration R-Net electrode system for EEG measurement. The experiment was created in SR Research's Experiment Builder, which allows connection to the above-mentioned EEG device. For recording stimulus onset in EEG data Trigger Box (BrainProducts) was used. EEG data were evaluated in BrainVision Analyzer 2 software.

The design of the experiment follows the Telpaz et al. (2015) methodology which was used, for example, also by Hu et al. (2020). Due to the need to measure ERP components repeatedly, the presentation of the stimuli was divided into 5 blocks, with 42 trials in each block, for a total of 210 trials (stimulus presentation). Trials were divided into the blank screen; drift correction (1000 ms); stimulus (2000 ms); blank screen. Between blocks, the participant had time to rest.

AOIs were created for the discount and review factors to quantify the eye-tracking data.

3 Results and Discussion

To evaluate early automatic attention Donoghue and Voytek (2022), Lin et al. (2018) the P200 component was analyzed. After visual inspection of the topographic and based on a previous study (Ma et al., 2018), electrodes Fz, F3, F4, Cz, C3, C4, P3, P4, and Pz were

---

[1] https://etlab.cz/

selected for deeper analysis in a time window of 160-210 ms after stimulus onset. The results showed a higher P200 amplitude in the case of hedonic products. In the case of the presentation of a hedonic product with a review, the amplitude decreased significantly. In the case of utility products, on the other hand, there was a longer latency, both when the product was presented separately and also with discounts and reviews.

The LPP component was analyzed in time 500-700 ms. Based on visual inspection, electrodes C3, C4, CP1, CP2, CP5, CP6, Cz, P3, P4, P7, P8 and Pz were then pooled.

Time to First Fixation, Duration of First Fixation, and Dwell Time in AOI metrics were chosen to evaluate the eye-tracking data.

4 Conclusion

The results showed that the use of the BrainProducts device for investigating consumer perception is real. Although 9 participants were a sufficient number to obtain smoothed ERP components, additional participants will likely be needed to obtain more generalizable results.

Acknowledgements

Literature

Donoghue, T., & Voytek, B. (2022). Automated meta-analysis of the event-related potential (ERP) literature. *Scientific Reports*, *12*(1), 1867.

Herbas T. B., Bjoern F., Takao, E. (2011). The Role of Hedonic and Utilitarian Attitudes in the Formation of Affective and Cognitive Customer Satisfaction. *95th Conference of the Japanese Society for Quality Control.*

Hu, F., Wu, Q., Li, Y., Xu, W., Zhao, L., & Sun, Q. (2020). Love at First Glance but Not After Deep Consideration: The Impact of Sexually Appealing Advertising on Product Preferences. *Frontiers in Neuroscience, 14*, 465.

Lin, M.-H. (J.), Cross, S. N.N. & Childers, T. L. (2018b). Understanding olfaction and emotions and the moderating role of individual differences. *European Journal of Marketing*, 52(3/4), 811-836.

Lin, M.-H. (Jenny), Cross, S. N. N., Jones, W. J., & Childers, T. L. (2018a). Applying EEG in consumer neuroscience. *European Journal of Marketing*, *52*(1/2), 66–91.

Ma, Y., Jin, J., Yu, W., Zhang, W., Xu, Z., & Ma, Q. (2018). How Is the Neural Response to the Design of Experience Goods Related to Personalized Preference? An Implicit View. *Frontiers in Neuroscience, 12*, 760.

Mastria, S., Ferrari, V., Codispoti, M. (2017). Emotional Picture Perception: Repetition Effects in Free-Viewing and during an Explicit Categorization Task. *Frontiers in Psychology*, *8*, 1001.

Pozharliev, R., Verbeke, W. J.M.I., Van Strien, J. W., Bagozzi, R. P. (2015). Merely Being with you Increases My Attention to Luxury Products: Using EEG to Understand Consumers' Emotional Experience with Luxury Branded Products. *Journal of Marketing Research*, *52*(4), 546-558.

Telpaz, A., Webb, R., & Levy, D. J. (2015). Using EEG to predict consumers' future choices. *Journal of Marketing Research, 52*(4), 511–529.

Spatial Geometry Imagination – Analysis of Problem-solving Strategies

Jana Slezáková[1], Stanislav Popelka[2], Roman Chvátal[3]

[1]Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, 17. listopadu 1192/12
779 00 Olomouc, Czech Republic; jana.slezakova@upol.cz
[2]Department of Geoinformatics, Faculty of Science, Palacký University Olomouc, 17. listopadu 50
771 46 Olomouc, Czech Republic; stanislav.popelka@upol.cz
[3]Department of Experimental Physics, Faculty of Science, Palacký University Olomouc, 17. listopadu 1192/12
779 00 Olomouc, Czech Republic; roman.chvatal@upol.cz

Extended Abstract

Purpose of the study

Our research focused on identifying selected solution strategies using eye-tracking. We were also interested in the test subject's attention and ability to combine different shapes. A non-standardized didactic test was used to detect geometric imagination in the plane (Slezáková, 2022). That test promotes the development of visual memory and allows the development of the ability to determine a new representation of an object after its transformation in the plane. The geometric imagination test contained a total of twelve images. Students worked only with pictures that consisted of non-convex polygons. The task of the tested pupil was to find out between which two (numbered) points a cut had to be made in order to form a regular hexagon when the two parts (in imagination only) were put together.

By setting the research questions, we tried to reveal the geometry problem-solving strategies and to add four research hypotheses:
H1: The most significant visual attention will be recorded when solving the last geometry problem in order.
H2: The solving strategies of the geometry problem without computation and the success rate of solving the geometry problem are correlated.
H3: The solving strategies of the geometry problem without computation and the speed of solving the geometry problem are correlated.
H4: The success rate of solving the geometry problem is not dependent on gender.

The research investigation took place in the Fort Science at Palacký University in Olomouc in the summer of 2021, at a time when intensive one-week suburban camps were organized. This was the period just after the end of the epidemic restrictions. We were thus able to provide a group of pupils of different ages, genders, and school types. A total of 53 pupils aged 11 to 14 participated in our research.

Research Methodology

The pupils' eye movements were recorded using a non-contact eye-tracker GazePoint GP3 HD with a sampling rate of 150 Hz. The eye-tracker was placed under a 24" IPS monitor, on which the didactic test was displayed in 1920 x 1080 px resolution.

During our investigation, we elucidated methodologies for addressing geometry problems devoid of computational approaches. Concurrently, we identified a correlation between the selected problem-solving strategy and the efficacy of solving geometry problems without resorting to computation. For a detailed analysis of each task's solution strategy, the fixation sequences recorded in the regions of interest were used, and the similarity of these sequences was measured using the ScanGraph tool (Dolezalova & Popelka, 2016; Popelka & Beitlova, 2022). The measured eye-tracking data were visually inspected in OGAMA (Voßkühler, Nordmeier, Kuchinke, & Jacobs, 2008). The eye movement trajectories of each respondent in the form of a video (GazeReplay function) and a static image (Scanpath function) were qualitatively assigned to one of three defined categories for each task.

Conclusions

The first hypothesis was confirmed. The highest visual attention was focused on the last task. Furthermore, we also confirmed the second hypothesis. The solving strategy and the geometry task's success rate are correlated.

In contrast, we expected the solution strategy and solution success to be correlated. However, our research did not confirm this. Furthermore, we hypothesized that there would be a relationship between the chosen solving strategy and solution speed. This was also not confirmed. The last thing was about the effect of gender on solution success rate, we did not confirm this either.

The study found that fewer fixations were required when the same problem was repeated, leading to faster solution times, as observed in tasks.

Our study underscored the imperative for further exploration of strategies for solving geometry problems without computational methods. Additionally, we plan to broaden our research scope by delving into the examination of metacognitive strategies applied in tackling tasks involving geometric imagination.

Literature

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques. *Journal of Eye Movement Research, 9*(4). https://doi.org/10.16910/jemr.9.4.5

Popelka, S., & Beitlova, M. (2022). *Scanpath Comparison using ScanGraph for Education and Learning Purposes: Summary of previous educational studies performed with the use of ScanGraph*. Paper presented at the 2022 Symposium on Eye Tracking Research and Applications, Seattle, WA, USA. https://doi.org/10.1145/3517031.3529243

Slezáková, J. (2022). *Trojúhelníkové figury: otestuj si geometrickou představivost.* UP Olomouc.

Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior research methods, 40*(4), 1150-1162. https://doi.org/10.3758/BRM.40.4.1150

Enhancing scientific reasoning and trust in science: the impact of a methodology course on unfounded beliefs – a pilot study

Viktória Sunyík

Institute of Experimental Psychology, Centre of Social and Psychological Sciences, SAS
Dúbravská cesta 9, 841 04 Bratislava, Slovakia
viktoria.sunyik@savba.sk

Abstract

We examined the effectiveness of a long-term intervention - a methodology course to enhance scientific reasoning and trust in science while reducing health-related epistemically unfounded beliefs (paranormal, conspiracy and pseudoscience) in a sample of university students (N = 142). A quasi-experimental pretest-posttest design was used. Although correlational analysis demonstrated negative relationships of scientific reasoning and trust in science with epistemically unfounded beliefs, a causal link was not observed. These results highlight the challenge of mitigating epistemically unfounded beliefs, even with a long-term intervention focused on teaching the principles of research methodology. While the expected intervention effect was not observed, this pilot study suggests potential positive impacts of the college experience itself on fostering better scientific reasoning. Additionally, it provides valuable insights for shaping future experimental research in this domain.

1 Introduction

The aim of the study was to examine the impact of long-term school intervention on increasing scientific reasoning and trust in science, thus reducing epistemically unfounded beliefs (EUB) i.e. beliefs not justified by the totality of available evidence and knowledge (Lobato et al., 2014). Scientific reasoning refers to specific skills needed for evaluating quality of scientific evidence (Drummond & Fischoff, 2017). Trust in science is a part of a broader attitudes toward science and refers to one's base-level tendency to either consider scientific methods, areas of study, experts, and the ideas they put forth positively or negatively (Hartman et al., 2017). Interventions aimed at improving scientific reasoning (for review see Engelmann et al., 2016) by focusing on the processes of scientific inquiry share some characteristics with interventions aimed at improving attitudes toward science as well (Lieskovský & Sunyík, 2022), and thus may be successful in improving both components simultaneously.

Such intervention aimed at deepening knowledge of the process of scientific inquiry could be covered by course of scientific methods. Previous research (e.g., Dyer & Hall, 2019; Stark, 2012; Vujaklija et al., 2010; Wilson, 2018) has indicated the effectiveness of completing methodology courses in enhancing scientific reasoning, improving attitudes towards science, and reducing certain types of EUBs. However, a notable research gap exists concerning the explicit examination of the collective influence of scientific reasoning and trust in science on mitigating unfounded beliefs. Therefore, our main assumption is that taking a methodology course, which has characteristics of an intervention promoting knowledge of the process of scientific inquiry, will develop scientific reasoning skills and also improve trust in science. At the same time, we expect a reduction in EUBs. Since our goal is to monitor the potential impact of scientific reasoning and trust in science, our intervention – the methodology course will not include explicit debunking of various types of EUBs, and thus we will investigate whether the assumed increase in scientific reasoning and trust in science alone will be sufficient for reducing health-related EUB.

2 Methods

2.1 Design and participants

A quasi-experimental pretest – posttest design was used, and final sample consisted of 142 university students majoring in psychology. Experimental group had 75 students ($M_{age} = 20.89$, $SD_{age} = 1.09$). Control group had 67 students ($M_{age} = 20.10$, $SD_{age} = 1.97$). We employed convenience sampling, as the students in the experimental group were second graders who were mandatory participants in the Methodology course, while the students in the control group were first graders without this course. Pretest data were collected at the beginning of the course in February 2022, and posttest data were collected at the end in May 2022.

2.2 The Methodology course as intervention for reducing unfounded beliefs

The content of the methodology course covered the fundamental aspects of scientific reasoning. It explained experimental, quasi-experimental and non-experimental designs, data collection methods, reliability and validity of psychological tests. The course combined the lecture part with the active participation of students, who were supposed to design their own research independently, defining the basic components: research problem, variables, hypotheses, questions, research type and procedure, and data collection methods. Such student activity is one of the main aspects of interventions to increase scientific reasoning as well as improving attitudes towards science (Lieskovský & Sunyík, 2022).

2.3 Materials

The study was a part of the dissertation thesis and here we focus on key variables relevant to the main aim of this paper. In the dissertation, we measured additional variables not reported here. All materials used in dissertation thesis regarding this study are available at https://osf.io/r6uv5/.

With one exception, we used the same materials in pretest and posttest.

Scientific reasoning. A Slovak version of Scientific Reasoning Scale (SRS) was used in pretest (Bašnáková et al., 2021) adapted from original scale developed by Drummond and Fischhoff (2017). Slovak version contains six short scenarios tapping some methodological principle such understanding the difference between correlation and causation. Participants agree or disagree with the final statement regarding each scenario. In posttest we used another version of SRS items (Bašnáková & Čavojová, 2019) to avoid effect of repeated measurement as this scale has true/false answer scoring format. Higher number of correct answers represent higher scientific reasoning skills.

Trust in science. We used a Credibility of Science Scale (Hartman et al., 2017) which contains 6 items. Participants indicated their agreement on a 7-point scale from 1 *(disagree very strongly)* to 7 *(agree very strongly)*. All items are reverse coded, higher score indicate higher trust in science.

Paranormal beliefs. We used General Magical Beliefs subscale from The Magical Beliefs About Food and Health Scale (Lindeman et al., 2000). Participants indicated their agreement with 10 items on a 5-point scale from 1 *(strongly disagree)* to 5 *(strongly agree)*. Higher score indicated higher acceptance of magical (paranormal) beliefs.

Pseudoscientific beliefs. We used three tools to measure this kind of unfounded beliefs: (1) A short 8-item version of Pseudoscientific belief scale (Fasce et al., 2021). Participants indicated their agreement on a 5-point scale from 1 *(strongly disagree)* to 5 *(strongly agree)*. Higher score indicated higher acceptance of pseudoscientific beliefs.

(2) The Cancer Awareness Measure of Mythical Causes Scale (Smith et al., 2018) to measure common cancer myths. Participants indicated their agreement with 12 items on a 5-point scale from 1 *(strongly disagree)* to 5 *(strongly agree)*. Higher score indicated higher acceptance of myths about causes of cancer.

(3) Complementary and Alternative Medicine (CAM) Subscale from Holistic Complementary and Alternative Medicine Questionnaire (Hyland et al., 2003). Participants indicated their agreement with 6 items on a 5-point scale from 1 *(strongly disagree)* to 5 *(strongly agree)*. Higher score indicated more positive attitudes toward CAM.

Conspiracy beliefs. We used the COVID-19 conspiracy subscale from the COVID-19 Unfounded Beliefs Scale (Teličák & Halama, 2022). Participants indicated their agreement with 6 items on a 5-points scale from 1 *(strongly disagree)* to 5 *(strongly agree)*. Higher score indicated higher acceptance of conspiracy beliefs.

3 Results & Discussion

Table 1 represent reliability coefficients and descriptive statistics of variables in pretest and posttest for both groups.

Tab. 1: Descriptive statistics in pretest and posttest

Variable		α	ω	Experimental M	Experimental SD	Control M	Control SD
Scientific reasoning	Pretest	.35	.39	5.03	0.87	4.42	1.35
	Posttest	.34	.37	4.29	1.23	3.60	1.21
Trust in science	Pretest	.85	.85	5.02	1.29	4.70	1.26
	Posttest	.90	.90	4.91	1.38	4.47	1.27
Magical beliefs	Pretest	.77	.78	2.73	0.67	2.94	0.63
	Posttest	.81	.81	2.66	0.74	2.97	0.64
COVID-19 conspiracy beliefs	Pretest	.95	.95	1.94	1.07	1.93	0.94
	Posttest	.95	.96	1.84	1.02	2.01	1.01
Pseudoscientific beliefs	Pretest	.64	.64	2.99	0.57	3.22	0.55
	Posttest	.64	.65	2.98	0.61	3.11	0.52
CAM beliefs	Pretest	.76	.76	2.52	0.77	2.93	0.70
	Posttest	.74	.75	2.50	0.80	2.79	0.71
Cancer myths beliefs	Pretest	.82	.83	2.94	0.62	3.02	0.52
	Posttest	.86	.87	2.90	0.70	3.14	0.54

α = Cronbach's alpha coefficient; ω = McDonald's omega coefficient; M = mean; SD = standard deviation.

Table 2 represent correlation among variable in pretest for both groups.

Tab. 2: Correlation in pretest for both groups

PRETEST	1.	2.	3.	4.	5.	6.	7.
1.Scientific reasoning	—	**.25***	-.12	-.23	-.10	-.16	-.05
2.Trust in science	**.24***	—	**-.43***	**-.47***	**-.33**	**-.32**	.09
3.Magical beliefs	**-.31**	**-.36**	—	**.29***	**.40**	**.45***	**.33**
4.C-19 conspiracy beliefs	**-.40***	**-.53***	**.33**	—	**.28***	.18	-.06
5.Pseudoscientific beliefs	-.19	**-.32**	**.57***	**.26***	—	**.25***	**.28***
6.CAM beliefs	-.10	-.19	**.36**	**.25***	**.37**	—	**.25***
7.Cancer myths beliefs	**-.25***	**-.25***	**.38**	**.51***	**.25***	.21	—

* p < .05, ** p < .01, *** p < .001. Pearson's coefficients are presented.
Below the diagonal = experimental group. Above the diagonal = control group.

Table 3 represent correlation among variable in posttest for both groups.

Tab. 3: Correlation in posttest for both group

POSTTEST	1.	2.	3.	4.	5.	6.	7.
1.Scientific reasoning	—	**.27***	**-.28***	**-.37**	-.16	-.18	-.02
2.Trust in science	**.33**	—	**-.37**	**-.59***	**-.40**	**-.25***	.05
3.Magical beliefs	**-.30**	**-.49***	—	**.25***	**.72***	**.49***	**.37**
4.C-19 conspiracy beliefs	**-.31**	**-.61***	**.42***	—	**.29***	**.34**	-.05
5.Pseudoscientific beliefs	**-.31**	**-.41***	**.69***	**.34**	—	**.47***	.18
6.CAM beliefs	-.14	**-.46***	**.49***	**.50***	**.46***	—	.18
7.Cancer myths beliefs	**-.35**	-.22	**.59***	**.41***	**.57***	**.33**	—

* p < .05, ** p < .01, *** p < .001. Pearson's coefficients are presented.
Below the diagonal = experimental group. Above the diagonal = control group.

Tables 2 and 3 display detailed correlations among all variables, with some exceptions, revealing expected patterns such as negative relationships between scientific reasoning and EUBs, as well as their negative associations with trust in science. These findings align with previous research results (Čavojová et al., 2020; Fasce & Picó, 2019; Sunyík & Čavojová, 2023). We identified a recurring positive relationship between scientific reasoning and trust in science, similarly like previous findings (Bašnáková et al., 2021; Sunyík & Čavojová, 2023) and positive relationships were observed between various types of EUBs, aligning with previous studies (Lobato et al., 2014).

Table 4 represent the results of ANCOVA model. The pretest data were used as covariates. Predictor variable is an enrollment in intervention (yes/no) thus "group" and the dependent variables are posttest data for scientific reasoning, trust in science and EUB.

Tab. 4: ANCOVA model testing the effect of intervention

		MS	F	p	η_p^2
Scientific reasoning	Group	6.33	5.00	**.027**	0.035
	Pretest	31.66	25.00	**<.001**	0.152
Trust in science	Group	1.48	1.79	.183	0.013
	Pretest	131.43	159.26	<.001	0.534
Magical beliefs	Group	0.81	3.24	.074	0.023
	Pretest	32.65	129.88	<.001	0.483
COVID-19 conspiracy beliefs	Group	1.13	2.82	.095	0.020
	Pretest	88.04	218.88	<.001	0.612
Pseudoscientific beliefs	Group	0.01	0.05	.826	0.000
	Pretest	11.57	48.11	<.001	0.257
CAM beliefs	Group	0.02	0.05	.820	0.000
	Pretest	33.29	98.72	<.001	0.415
Cancer myths beliefs	Group	1.25	5.78	**.017**	0.040
	Pretest	24.91	115.29	**<.001**	0.453

MS = Mean square; F = Test statistic; p – significance; η_p^2 = partial eta squared effect size.

Although comparison of posttest scores between both groups, taking pretest scores into account, revealed a significant effect of the methodology course on scientific reasoning and cancer myth beliefs, effects are very small to negligible. Moreover, when considering pretest and posttest scores of scientific reasoning of both groups, we see their decline, what oppose our main assumption. Change in cancer myths beliefs was caused by a statistically significant increase in the score on this variable in the control group ($M_{pretest}$ = 3.02, $M_{posttest}$ = 3.14, t (66) = -2.53, p = .014). There was no significant reduction of this type of EUB in the experimental group. Our results did not support the primary assumptions regarding the impact of the methodology course on enhancing scientific reasoning, trust in science, and reducing unfounded beliefs. Possible explanations may lie in methodological aspects of this study - we did not use random sampling, but convenience and our group were not equivalent, and we found lower values of reliability in case of scientific reasoning scale. At the start of the quasi-experiment, students in the experimental group showed better scientific reasoning (t (110) = 3.15, p = .002) and fewer pseudoscientific beliefs (t (139) = -2.39, p = .018) and fewer pro-CAM attitudes (t (140) = -3.31, p = .001) compared to the control group, who were a year younger. This suggests that even one year of university education has the potential to foster skepticism towards unfounded beliefs and affect scientific reasoning. Our findings imply that it's not merely completing one course, but rather the entire university experience up to that point, that may contribute to this shift. On the other hand, another confounding variable may played role such characteristics of the teacher (Lieskovský & Sunyík, 2022) and the difficulty of the methodology course itself

(Onwuegbuzie et al., 2000). Future research should explore these potential factors further.

4   Conclusion

In this pilot study we examined the effect of a long-term intervention to enhance scientific reasoning and trust in science while reducing epistemically unfounded beliefs. Although results did not support our assumptions, they underscore the challenges associated with reducing unfounded beliefs and conducting field experiments. Moreover, they emphasize the importance of considering students' university experience itself as a possible factor in reducing unfounded beliefs and improving scientific reasoning. In this context, maturation should also be considered as a potential confounding variable. Additionally, other potential confounders such as the characteristics of the person providing the intervention, its difficulty, and the duration of the intervention should be considered.

Acknowledgements

References

Bašnáková, J. & Čavojová, V. (2019). Sú vedci „vedecky gramotní"? Čo je vedecká gramotnosť a ako ju merať. In Piterová, I., & Výrost, J. (Eds.) Sociálne procesy a osobnosť 2018. Zborník príspevkov z medzinárodnej konferencie, (s. 50-57). Košice: Spoločenskovedný ústav CSPV SAV, Košice.

Bašnáková, J., Čavojová, V., & Šrol, J. (2021). Does Concrete Content Help People to Reason Scientifically?: Adaptation of Scientific Reasoning Scale. *Science & Education*, *30*(4), 809–826. https://doi.org/10.1007/s11191-021-00207-0

Čavojová, V., Šrol, J., & Jurkovič, M. (2020). Why should we try to think like scientists? Scientific reasoning and susceptibility to epistemically suspect beliefs and cognitive biases. *Applied Cognitive Psychology*, *34*(1), 85–95. https://doi.org/10.1002/acp.3595

Drummond, C., & Fischhoff, B. (2017). Development and Validation of the Scientific Reasoning Scale.

Journal of Behavioral Decision Making, *30*(1), 26–38. https://doi.org/10.1002/bdm.1906

Dyer, K. D., & Hall, R. E. (2019). Effect of Critical Thinking Education on Epistemically Unwarranted Beliefs in College Students. *Research in Higher Education*, *60*(3), 293–314. https://doi.org/10.1007/s11162-018-9513-3

Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education – meta-analytic evidence from intervention studies. Educational Research and Evaluation, 22(5-6), 333–349. https://doi.org/10.1080/13803611.2016.1240089

Fasce, A., Avendaño, D., & Adrián-Ventura, J. (2021). Revised and short versions of the pseudoscientific belief scale. *Applied Cognitive Psychology*, *35*(3), 828–832. https://doi.org/10.1002/acp.3811

Fasce, A., & Picó, A. (2019). Science as a Vaccine: The Relation between Scientific Literacy and Unwarranted Beliefs. *Science & Education*, *28*(1–2), 109–125. https://doi.org/10.1007/s11191-018-00022-0

Hartman, R. O., Dieckmann, N. F., Sprenger, A. M., Stastny, B. J., & DeMarree, K. G. (2017). Modeling Attitudes Toward Science: Development and Validation of the Credibility of Science Scale. *Basic and Applied Social Psychology*, *39*(6), 358–371. https://doi.org/10.1080/01973533.2017.1372284

Hyland, M. E., Lewith, G. T., & Westoby, C. (2003). Developing a measure of attitudes: The holistic complementary and alternative medicine questionnaire. *Complementary Therapies in Medicine*, *11*(1), 33–38. https://doi.org/10.1016/S0965-2299(02)00113-9

Lieskovský, J., & Sunyík, V. (2022). How to enhance scientific literacy? Review of interventions focused on improving high school students' scientific reasoning skills and attitudes toward science. *Československá Psychologie*, *66*(1), Article 1. https://doi.org/10.51561/cspsych.66.1.30

Lindeman, M., Keskivaara, P., & Roschier, M. (2000). Assessment of Magical Beliefs about Food and Health. *Journal of Health Psychology*, *5*(2), 195–209. https://doi.org/10.1177/135910530000500210

Lobato, E., Mendoza, J., Sims, V., & Chin, M. (2014). Examining the Relationship Between Conspiracy Theories, Paranormal Beliefs, and Pseudoscience Acceptance Among a University Population: Relationship between unwarranted beliefs. *Applied Cognitive Psychology*, *28*(5), 617–625. https://doi.org/10.1002/acp.3042

Onwuegbuzie, A. J., Slate, J. R., Paterson, F. R. A., Watson, M. H., & Schwartz, R. A. (2000). Factors Associated with Achievement in Educational Research Courses. *Research in the Schools*, *7*(1), 53–65.

Smith, S. G., Beard, E., McGowan, J. A., Fox, E., Cook, C., Pal, R., Waller, J., & Shahab, L. (2018). Development of a tool to assess beliefs about mythical causes of cancer: The Cancer Awareness Measure Mythical Causes Scale. *BMJ Open*, *8*(12), e022825. https://doi.org/10.1136/bmjopen-2018-022825

Stark, E. (2012). Enhancing and Assessing Critical Thinking in a Psychological Research Methods Course. *Teaching of Psychology, 39* (2), 107-112. https://doi.org/10.1177/0098628312437725

Sunyík, V., & Cavojova, V. (2023). Alternative Medicine, COVID-19 Conspiracies, and Other Health-Related Unfounded Beliefs: The Role of Scientific Literacy, Analytical Thinking, and Importance of Epistemic Rationality. *Studia Psychologica*, *65*, 246–261. https://doi.org/10.31577/sp.2023.03.878

Teličák, P. & Halama, P. (2022). Konštrukcia a psychometrická analýza vlastností škály Covid-19 nepodložených presvedčení. In Čavojová,V. & Halama, P. eds. Prežívanie a dôsledky pandémie COVID-19 na Slovensku: Pohľad sociálnych vied. (37-44). Ústav experimentálnej psychológie, Centrum spoločenských a psychologických vied SAV, v.v.i. Bratislava

Vujaklija, A., Hren, D., Sambunjak, D., Vodopivec, I., Ivaniš, A., Marušić, A., & Marušić, M. (2010). Can teaching research methodology influence students' attitude toward science? Cohort study and nonrandomized trial in a single medical school. Journal of Investigative Medicine, 58(2), 282-286

Wilson, J.A. (2018). Reducing pseudoscientific and paranormal beliefs in university students through a course in science and critical thinking. Science & Education, 27, 183–210. https://doi.org/10.1007/s11191-018-9956-0

Auxiliary unsupervised loss for supervised learning

Sabína Samporová a Kristína Malinovská

Katedra aplikovanej informatiky, FMFI,
Univerzita Komenského v Bratislave
Mlynská dolina, 84248 Bratislava
Email: samporova1, rebrova1 @uniba.sk

Abstrakt

We present the design of a novel learning method for improved neural network learning that takes into account the data similarities determined by a self-organizing network. We incorporate self-organization as a part of the loss function and show the use of our concept in a new neural network for supervised learning. We compare how the success rate of a classical perceptron increases with this added component of the loss in a classification task. Our results show that with a minimalist model architecture in terms of the hidden layer size, our model convergence is more reliably, compared to the classical multi-layer perceptron. Even the classification accuracy of our model is higher. Our model achieves mean accuracy 89%, while the MLP only 82%.

1 Introduction and motivation

Our overall goal is to improve the semi-supervised learning of deep neural networks. In this paradigm, we use a limited amount of labeled data, but we are also looking for ways to incorporate unlabeled data into the learning. Our improvement lies in the application of topological self-organization. Self-organizing models can learn from unlabeled data, allowing us to make use of all available data. We believe that the additional information from the unlabeled data will support more efficient model learning.

2 SOM

For our task of augmenting the supervised loss with unsupervised mechanism we selected the self-organizing map (SOM, Kohonen, 1990) as a suitable model for implementing self-organization over the input data. Using SOM, the data is nonlinearly transformed into a 2D map and divided into clusters according to the classes to which they belong. In a well-trained map, similar inputs are represented by data models, prototypes (winning neurons), that are close to each other, within the same cluster, therefore preserving the similarities in the topology of the network. Since this is an unsupervised learning model, there is no need to know the labels of the input data.
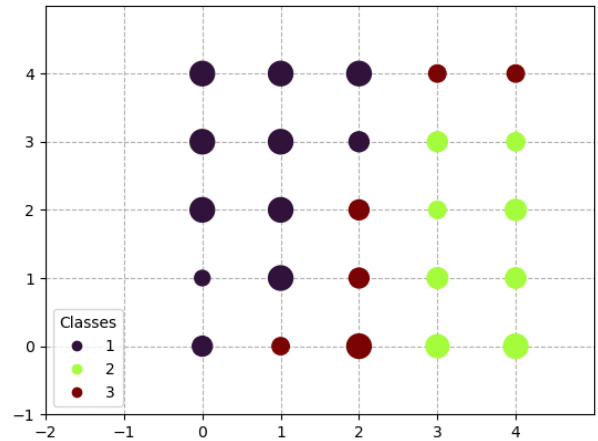

Fig. 1: Self-organizing map trained on Wine dataset

3 Loss function design

We present auxiliary SOM-based loss function J_S. The input of the function are triplets of input data points. Such an approach using triples of data inputs is well known in the field of deep neural networks and has been used primarily for the visual data classification task (Chechik et al., 2010). The first item of i-th triplet is a particular input from a class, which we denote by x_i. The second part is a different input also from the same class, denoted by z_i. The third part is an input that is from a different class than the other two, let us denote it as ξ_i.

We train the SOM using all the items from triplets. During model training, we compute the Euclidean distances D_E between the SOM prototypes (denoted as p) of pairs x_i, z_i and x_i, ξ_i, and combine these distances in one batch as in Eq. 1 and Eq. 2, denoted as D_C and D_I (congruent and incongruent). In a well-trained network, D_C should be small and D_I should be large. Therefore we compute their relative relationship, as given in Eq. 3.

$$D_C = \frac{1}{n} \sum_{i}^{n} D_E(p(x_i), p(z_i)) \qquad (1)$$
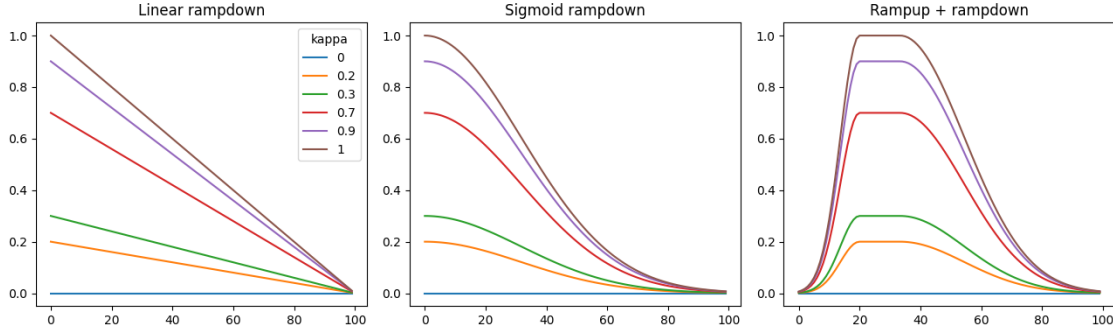
Fig. 2: Investigated ways of ramping the hyperparameter κ in time

$$D_I = \frac{1}{n} \sum_i^n D_E(p(x_i), p(\xi_i)) \qquad (2)$$

Subsequently we rescale this relation to the interval $[0, 1]$ according to the equation 4 and obtain the formula of the loss function whose value is close to zero if D_C is small and D_I is large. If D_C and D_I are different, the value of the loss function increases and approaches to the value 1.

$$D = \frac{D_I - D_C}{D_I + D_C} \qquad (3)$$

$$J_S(\tau) = \frac{1}{2} - \frac{1}{2} \cdot D = \frac{1}{2}\left(1 - \frac{D_I - D_C}{D_I + D_C}\right) = \qquad (4)$$

$$\frac{1}{2}\left(\frac{D_I + D_C}{D_I + D_C} - \frac{D_I - D_C}{D_I + D_C}\right) = \frac{D_C}{D_I + D_C}$$

4 MLP-SOM model

We propose the new MLP-SOM model that links the multi-layer perceptron (MLP) and the distance-based SOM loss function J_S from pre-trained SOM. The combined loss function (Eq. 6) is a combination of the supervised loss function S and proposed unsupervised loss J_S. The supervised loss is the mean squared error of the model prediction of the first part of the triplet, denoted $g(x_i, \theta)$ and the expected label $\hat{y}$. The loss function J_S is scaled by the parameter $\kappa(t)$, which is linearly ramped down. Its value decreases linearly from a maximum value equal to κ to a value of zero in time. We also take into consideration different ramping strategies such as the Sigmoid function. The $\kappa(t)$ denotes the value of the parameter used in the relation at time t. The maximum value of κ is a hyperparameter that we experimentally explore different values.

$$S(\theta) = \frac{1}{n} \sum_i^n ||g(x_i, \theta), \hat{y}||^2 \qquad (5)$$

$$Loss(\theta) = S(\theta) + \kappa(t) \cdot J_S(\tau) \qquad (6)$$

5 Experiments and results

We decided to validate the proposed MLP-SOM model experimentally on a multi-class classification task. The famous tabular dataset we used, Wine dataset, was created by Aeberhard & Forina (1991). We used the pre-trained SOM illustrated in Fig.1. Looking at the projection of this SOM we see the map is quite well organized with the individual clusters corresponding to the 3 classes of the learned dataset.

5.1 Wine dataset

The data in the Wine dataset were the result of chemical analysis of wines produced in the same region in Italy, from three different varieties. Each data sample is vector of 13 values. We divided the data into training (75%) and test (25%). Within the training set, we created triples from the data. Each sample was paired 25 times with sample from the same class and sample from a different class. In this way, we obtained a training set containing 6650 triples.

5.2 Compared models

In the experiment, we compared the accuracy of a multi-layer perceptron (MLP) baseline with our MLP-SOM model that uses the combined loss function. In both cases, the perceptrons had the same architecture, namely 13 neurons in the input layer, 15 neurons in the hidden layer, and 3 neurons in the output layer. The activation function for the neurons in the input and hidden layers was the Sigmoid and for the output layer we used the Softmax function. The models were implemented using the PyTorch library (Paszke et al., 2019) using the default weight initialization. Other hyperparameters were the learning rate set to 0.0001 and the type of optimizer, which was Adam.
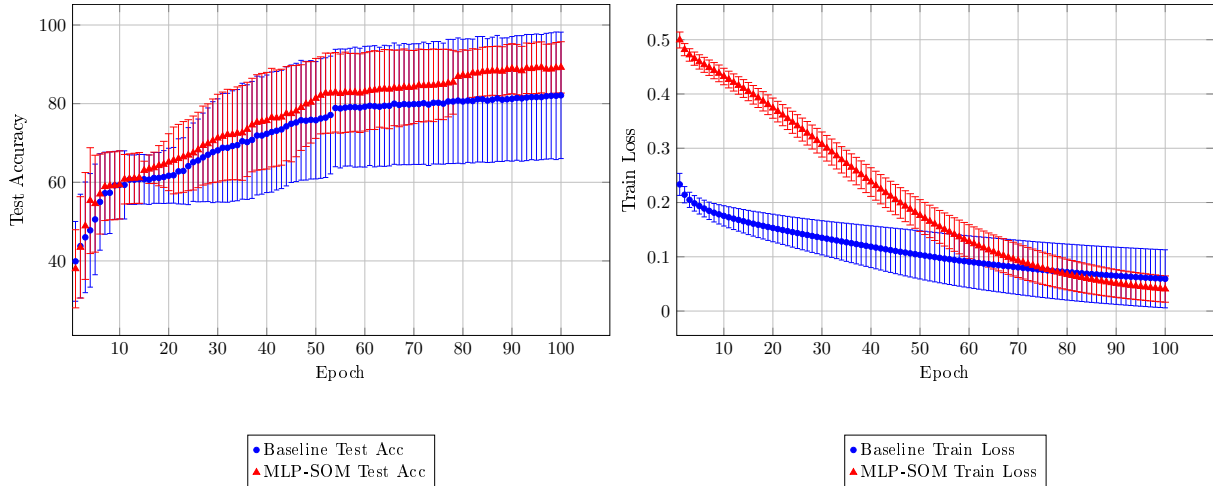
Fig. 3: Evolution of baseline and MLP-SOM model metrics during training (testing accuracy, training loss function)

6   Results of experiments

The results of experiments are the average performance of 20 models after a fixed amount of 100 training epochs. We experimented with different kinds of weight initialization and even smaller architectures. In all experiments, our best performing model was at least 2% better in accuracy than the baseline.

In the main experiment, we investigated different kinds of ramping of the κ and maximum value of κ. Ramps are shown in Fig. 2. The resulting test accuracies of the baseline and models with 5 different κ values are shown in table 1. In this experiment, we obtained the best result so far with the dataset, namely with the MLP-SOM model ($\kappa = 0.7$) and Sigmoid ramping. The ramping method combining rampup and rampdown was the least successful. For the best model, the test accuracy is $89.22\% \pm 6.53\%$, which is more than 7% better than the baseline model without using the SOM loss function. Fig.3 shows a comparison of the evolution of the two models during training.

κ	linear	sigmoid	up + down
0	82.11 ± 16.08		
0.2	84.00 ± 10.74	85.11 ± 14.73	80.89 ± 16.59
0.3	81.56 ± 13.84	85.33 ± 11.38	79.78 ± 18.20
0.7	83.44 ± 15.37	89.22 ± 06.53	79.89 ± 14.25
0.9	84.67 ± 10.40	77.89 ± 18.59	74.78 ± 19.17
1	82.67 ± 11.98	79.11 ± 16.61	78.44 ± 18.62

Tab. 1: Testing accuracy after 100 epochs. $\kappa = 0$ indicates the baseline MLP.

7   Conclusion

In our research, we have been developing and testing a method to improve the classification accuracy of a neural network trained with supervised learning. Our method was based on the use of an unsupervised neural model - a self-organizing map that is trained from data that does not need to have labels. We proposed an MLP-SOM model, using predictions from the pretrained self-organizing map as part of the loss function. We experimentally tested this model on the tabular Wine classification dataset. We investigated various hyperparameters and the results have shown that our model had better testing accuracy, compared to a model that did not use an additional SOM-based loss function. Given this result, we believe that even in the case of semi-supervised learning, where most of the data does not have an associated label, the SOM loss function can be a useful tool to improve the training and increase the classification success of the model.

References

Aeberhard, S. & Forina, M. (1991). Wine. UCI Machine Learning Repository.

Chechik, G., Sharma, V., Shalit, U. & Bengio, S. (2010). Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 11(3).

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. V *2015 IEEE In-*

---

[1]https://cogsci.fmph.uniba.sk/sskv/

ternational Conference on Computer Vision (ICCV), str. 1026–1034.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480.

Krizhevsky, A., Hinton, G. et al. (2009). Learning multiple layers of features from tiny images. Technická správa, University of Toronto, Toronto.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., De-Vito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. V *Advances in Neural Information Processing Systems 32*, str. 8024–8035. Curran Associates, Inc.

Tarvainen, A. & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. & Garnett, R. (zost.), V *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc.

The concept of the project:
Dyslexia re-education and development of reading skills utilising eye-tracking

Čeněk Šašinka, Nicol Dostálová, Stanislav Ježek, Jiřina Bednářová, Kateřina Jochecová, Petra Daňsová, Roman Švaříček, Stanislav Popelka, Michaela Vojtěchovská, David Košatka

Masaryk University, Brno

cenek.sasinka@mail.muni.cz, nicol.dostalova@mail.muni.cz, jezek@fss.muni.cz, jir.bednarova@gmail.com, katerina.jochecova@mail.muni.cz, 363837@mail.muni.cz, svaricek@phil.muni.cz, standa.popelka@gmail.com, kosatka@phil.muni.cz

Extended Abstract

The contribution aims to present the overall concept of the DeveLex project, which aims not only at research and development of the proposed solution but also its viable implementation in practice. The project's objective is to develop interactive automated adaptive software and intervention tasks for dyslexia re-education and the development of reading skills. The presented system, DeveLex, is based on the principle of eye tracking and, thanks to an advanced computational module (cognitive modelling), allows real-time adaptation to the current level of the child's abilities. The principle of the zone of proximal development is utilised. The project encompasses not only an IT solution and re-education tasks but also the adaptation and development of the entire system of care and services in the field of dyslexia intervention. The project aims to foster collaboration between Pedagogical Psychological Counselling Centre, the library network, and parents. It integrates knowledge from the fields of pedagogy, statistics, psychology, optometry, and IT.

Program **SIGMA**

A Model of the Reference Frame of the Ventriloquism Aftereffect Considering Auditory Saccade Adaptation

Peter Lokša, Norbert Kopčo

Institute of Computer Science, P. J. Šafárik University, Košice
Jesenná 5, 040 01 Košice, Slovakia
peter.loksa@upjs.sk, norbert.kopco@upjs.sk

Abstract

Background: The ventriloquism aftereffect (VAE), observed as a shift in the perceived locations of sounds after audio-visual stimulation, requires reference frame (RF) alignment since hearing and vision encode space in different frames (head-centered, HC vs. eye-centered, EC). Previous experimental studies observed inconsistent results: a mixture of head-centered and eye-centered frames for the VAE induced in the central region vs. a predominantly head-centered frame for the VAE induced in the periphery. A previous model proposed to describe these data required different parameter fits to predict the central vs. peripheral aftereffect data. Here, a new version of the model is introduced to provide a unified prediction of all data sets considering that saccade responses, used to measure VAE, are also adapted.

Methods: VAE was measured using eye-tracked saccades to the perceived locations of sound (i.e., "auditory saccades") presented via loudspeakers. The model has two components: an auditory space representation component and a saccade-representation component. The former is adapted by ventriloquism signals in the head-centered reference frame. The later one characterizes adaptation in auditory saccade responses in eye-centered frame.

Results: The updated version of the model provides a unified prediction of all the data, even if only head-centered RF is considered in the auditory space representation, while proposing that there are a priori biases in the auditory saccades and that the auditory saccades also undergo ventriloquism adaptation.

Conclusion: The results suggest that purely head-centered RF is used for adaptation of auditory spatial representation in the ventriloquism aftereffect, and that the apparently mixed eye-and-head centered RF observed experimentally is most probably due to adaptation in auditory saccades that are eye-centered.

1   Introduction

The neural representations of visual and auditory space use different reference frames. Vision is referenced relative to the direction of eye-gaze (eye-centered), while hearing is referenced relative to the head orientation (head-centered). The current study examines how are these two representations aligned at higher level of spatial processing to allow visually guided adaptation of auditory spatial perception.

Existing models of the audio-visual (AV) RF alignment only consider integration when in the auditory and visual stimuli are presented simultaneously (i.e., the ventriloquism effect; VE) (Razavi et al., 2007; Pouget et al., 2002). We proposed a model of the visually guided adaptation of auditory spatial representation in VAE (Lokša & Kopčo, 2023) to describe behavioral data of Kopčo et al. (2009, 2019). Here extensions of the model are introduced: to characterize the mixed RF of VAE observed in Kopčo et al. (2009); to provide a unified account of conflicting results of Kopčo et al. (2009, 2019)

In addition to auditory space representation in HC RF, the current model and the dHEC model (Lokša & Kopčo, 2023) model consider 3 candidate mechanisms underlying these effects: eye-centered signals influencing auditory space representation, fixation-position-dependent attenuation in auditory space adaptation, adaptation in the saccades used for responding in the experiments.

Finally, Kopco et al. (2019) observed a new adaptive phenomenon induced by aligned audiovisual stimuli presented in the periphery that is also considered.

2   Model

The SA-HC model (Fig. 1) focuses on explaining both central and peripheral data using one mechanism. It assumes that auditory space is adapted by visual signals only in HC RF (like in the basic version of dHEC model), while the saccades, used for responding, are also adapted – in EC RF. Specifically, it assumes that (1) during training, the saccades are adapted to be hypometric or hypermetric, depending on FP, A component, and V component locations, (2) during testing, the adapted saccades either enhance or reduce the bias due to auditory space representation, depending on the A target location vs FP.
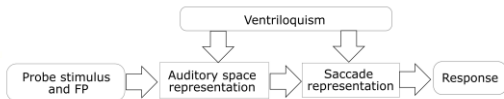
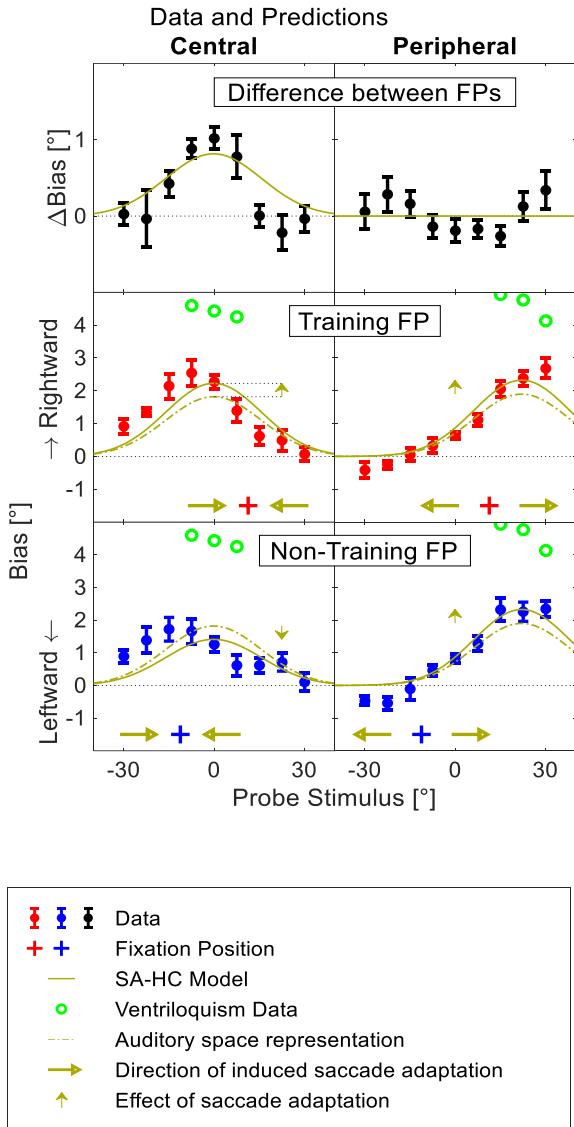Fig. 1: Diagram of the SA-HC model.

Data and Predictions

Fig. 2: SA-HC model evaluation on central & peripheral data. Model predictions (solid lines) and experimental data (symbols). Arrows and dash-dotted lines illustrate model operation: Ventriloquism (green circles) determines auditory space adaptation (dashed lines) which is additionally modified by the ventriloquism-adapted hypo/hypermetric saccades (horizontal arrows).

3   Results

SA-HC model fits the central and peripheral data simultaneously (Fig. 2). Since the model only uses HC RF to induce VAE, this result supports the conclusion

that reference frame of VAE is purely head-centered, and the previously observed mixed RF was due to saccade adaptation.

4   Conclusion

We introduced an extended model to describe the reference frame of ventriloquism aftereffect data of Kopčo et al. (2009, 2019).

A previous model by Lokša and Kopčo (2023) was able to explain central adaptation and peripheral adaptation results of Kopčo et al. (2009, 2019) when fitted to the data separately, i.e., with different values of model parameters. Thus, the inconsistency between the behavioral results were not reconciled.

The current model, incorporating auditory saccade adaptation, can explain the central and the peripheral data simultaneously which confirms that RF of VAE is most likely not mixed.

Next steps are to (1) explain ventriloquism-like adaptation induced by AV-aligned stimuli in Kopčo et al. (2019), and to (2) experimentally test the model predictions about saccade-related EC bias and saccade representation adaptation, as well as the prediction that the reference frame of the VAE is purely head-centered.

Acknowledgements

Literature

Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*, 257-262.

Knudsen, E. I. & Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science, 230*, 545-548.

Knudsen, E. I. & Knudsen, P. F. (1989) Vision calibrates sound localization in developing barn owls. *Journal of Neuroscience*, *9*, 3306-3313.

Kopčo, N., Lin, I., Shinn-Cunningham, B. G. & Groh, J. M., (2009). Reference Frame of the Ventriloquism Aftereffect. *Journal of Neuroscience*, *29*(44), 13809-13814.

Kopčo, N., Lokša, P., Lin, I., Groh, J. M., & Shinn-Cunningham, B. G. (2019). Hemisphere-specific properties of the ventriloquism aftereffect. JASA 146, EL177.

Maier, J. K., McAlpine, D., Klump, G. M., & Pressnitzer, D., (2009). Context Effect in the

Discriminability of Spatial Cues. *JARO 11*(2), 319-328.

Brainard, M. S., & Knudsen, E. I. (1995). Dynamics of visually guided auditory plasticity in the optic tectum of the barn owl. *J Neurophysiol, 73*, 595–614.

Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nat Rev Neurosci, 3*(9), 741-7. https://doi.org/10.1038/nrn914

Razavi, B., O'Neill, W.E., & Paige, G.D. (2007). Auditory spatial perception dynamically realigns with changing eye position. *J Neurosci*, *27*, 10249–10258.

Harris, C. M. (1994). Does saccadic undershoot minimize saccadic flight-time? A Monte-Carlo study. *Vision Res*, *35*(5), 691-701.

Lokša, P. & Kopčo, N. (2016). Modelling of the Reference Frame of the Ventriloquism Aftereffect. *Sborník z 16.ročníku konference Kognice a Umělý život*. 101-106.

Yao, L. J. & Peck, C. K. (1997). *Saccadic eye movements to visual and auditory targets*. Springer, 25-34.

Lokša, P. & Kopčo, N. (2023). Toward a Unified Theory of the Reference Frame of the Ventriloquism Aftereffect. *Trends in Hearing*, *27*.

Workshop: Possibilities of Eye-tracking Data Visualization and Analyses

Stanislav Popelka, Kamila Fačevicová, Michaela Vojtěchovská

Palacký University Olomouc
17. listopadu 50, 77146, Olomouc, Czech Republic
stanislav.popelka@upol.cz

Abstract

This workshop is tailored to provide attendees with an in-depth understanding of eye-tracking technology and its capacity for analyzing behavior in response to various visual stimuli. It will delve into the foundational aspects of eye-tracking technology, encompassing its historical evolution, core principles, advantages, and limitations. The focus will extend to how this technology can unveil the cognitive processes engaged when individuals interact with visual materials.

Beyond introductory content, the session will feature practical demonstrations of case studies in cartography, repurposed here as examples applicable across different visual contexts. A key part of the workshop will be dedicated to introducing methodologies, strategies, and tools developed and employed at Palacký University in Olomouc.

The workshop begins with an introduction to GazePlotter and ScanGraph, two advanced tools that exemplify the practical application of eye-tracking data. **GazePlotter** focuses on visualizing attention distribution across different Areas of Interest, highlighting where participants look first, how long they gaze, and the pattern of their visual journey. **ScanGraph** complements this by offering scanpath comparison techniques, enabling an analysis of the sequence in which Areas of Interest are visited and how this sequence varies across individuals or stimuli.

Following this, the workshop will introduce the **Compositional Data Analysis** (CoDA) method, emphasizing its unique approach to handling data that are inherently relative, using a log-ratio transformation for metrics such as the number and duration of visual fixations.

A significant addition to this suite of tools is eyetRack, a new R package and a Shiny application tailored for the accessible analysis of eye-tracking data from SMI or Tobii eye-trackers. eyetRack simplifies the initial exploration of eye-tracking data, focusing on the number and duration of fixations. What sets eyetRack apart is its use of recurrence and recurrence quantification analysis (RQA) for a more nuanced analysis. These techniques not only allow for the visualization of recurrent fixations but also employ RQA measures to quantify these observations, reducing the likelihood of subjective bias in result interpretation.

Furthermore, eyetRack calculates the coefficient K, distinguishing between focal and ambient attention, adding depth to the understanding of how viewers interact with visual stimuli.

Participants in this workshop will acquire not just theoretical knowledge but also practical skills in utilizing eye-tracking technology across a variety of applications. Through demonstrations and hands-on examples involving GazePlotter, ScanGraph, CoDA, and eyetRack, attendees will explore different methods, approaches, and tools that offer insights into visual cognition. This knowledge is invaluable for professionals seeking to enhance visual communication and design more effective and engaging visual presentations.

Acknowledgements

Workshop: Eye Tracking in Virtual reality - Pico Neo III Pro Eye Implementation

Ondřej Kvarda[1], Kateřina Jochecová[2] , Gabriela Godišková[1]

[1]Department of Geography, Faculty of Science, Masaryk University, Brno, Czech Republic
[2]Department of Information and Library Studies, Faculty of Arts, Masaryk University, Brno, Czech Republic
kvarda.ondrej@mail.muni.cz, katerina.jochecova@mail.muni.cz

Abstract

With the advent of standalone VR HMDs, eye tracking (ET) implementations quickly followed. Currently, they are mostly created to heighten the gaming experience but can also serve as an advantageous research tool. The most common approach to measuring eye movements in HMDs is video oculography (VOG), and it's employed by the biggest eye tracker providers in the market, including Tobii, Pupil Labs, or Varjo. However, eye tracking data quality and the calibration itself remain one of the biggest challenges in ET in HMDs, as the precision, accuracy, and other metrics are often insufficient or not reported by the manufacturer at all (Adhanom et al., 2023). Researchers are currently trying to get more credible results by implementing their own solutions.

The presented solution is a custom implementation of eye tracking calibration and logging functionality for the Pico Neo III Pro Eye standalone VR HMD with the use of the Tobii XR SDK in an immersive virtual environment created in the Unity game engine. Typically, to get raw ET data from the Tobii XR SDK, one would need to purchase a proprietary Tobii Ocumen Studio. However, basic ET data can be obtained from the base SDK as well. With the use of the "Gaze Visualizer" object from the Tobii XR SDK and a Raycast functionality from Unity Scripting API, we are able to devise a simple script that logs the exact coordinates of the user and his gaze in a virtual environment throughout time while also simultaneously logging the names of particular objects the user is looking at. The frequency of this logging can be modified and can be set as low as approximately 10-15 ms as it is limited by the ET implementation that has a refresh rate of 90Hz. From these data, saccades, fixations, dwell times, gaze heat maps, and gaze paths can be calculated in post-processing. Calibration of ET can be done in a separate Tobii calibration app native to the Pico Neo III Pro Eye. Nevertheless, no raw data can be extracted from this app. For the purpose of objective ET calibration, a separate scene in our app was created where the user is tasked with focusing on appearing targets spread around the environment and then the average distance between the user's gaze and the center of each target in time is calculated; the average distance is then taken as an indicator of ET (in)accuracy. The ET data is logged and saved in the internal memory of the VR HMD in the form of CSV files and thus it is easily accessible and evaluable.

This implementation is applied in a custom solution focused on testing the effect of perspective and embodiment on the user's perception and evaluation of cartographic tasks involving altitude, line of sight, and route length, as well as in a solution focused on different cartographic visualization methods for bivariate data in immersive virtual environments.

Acknowledgements

Literature

Adhanom, I.B., MacNeilage, P. & Folmer, E. (2023). Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. *Virtual Reality, 27*. https://doi.org/10.1007/s10055-022-00738-z

Workshop: Introduction to the current state of dyslexia intervention, the possibility of eye-tracking technology and the needs of target users

Čeněk Šašinka[1], Nicol Dostálová[2], Michaela Vojtěchovská[3], Kateřina Jochecová[1], David Košatka[1], Stanislav Ježek[4], Petra Pátková Daňsová[4]

[1]Department of Information and Library Studies, Faculty of Arts, Masaryk University, Brno, 602 00, Czech Republic

[2]Department of Educational Sciences, Faculty of Arts, Masaryk University, Brno, 602 00 Czech Republic

[3]Department of Geoinformatics, Palacký University Olomouc, 17. listopadu 710/50, 779 00, Czech Republic

[4]Department of Psychology, Faculty of Social Sciences, Masaryk University, Brno, Joštova 218, 602 00, Czech Republic

cenek.sasinka@mail.muni.cz, nicol.dostalova@mail.muni.cz, mail@vojtechovska.com, katerina.jochecova@mail.muni.cz, kosatka@phil.muni.cz, jezek@fss.muni.cz, 363837@mail.muni.cz

Extended abstract

Dyslexia is a specific learning disorder manifested by reading difficulties (Lyon et al., 2003). It is important to focus on precise and accurate diagnostics of dyslexia followed by a complex and regular intervention method that would encourage the child dyslexic readers to develop and strengthen their reading skills. For these purposes, we are aiming to develop a complex system that would support child dyslexic readers in their overall reading development and enhancement of their reading skills.

The main goal of the workshop is to introduce the conceptualization of a comprehensive system called DeveLex that aims at a complex dyslexia intervention process based on 3 crucial parameters: adaptability, scalability, and interactivity to support the dyslexic reading performance taking into account the needs and variability of dyslexic child reading progress. The presented DeveLex system is based on the principle of tracking eye movements to monitor the overall reading processes of an individual, establishing a proper user interface, and also to allow a real-time adaptation to the current level of the child's abilities using the principle of the zone of proximal development in combination with advanced cognitive modeling of the reading processes. The DeveLex system will include a sophisticated IT solution, a battery of tasks (e.g. paired reading and more) for the dyslexia intervention and the development of reading skills, and also the adaptation and development of the entire system of intervention and care services for dyslexic individuals.

The project also aims to create collaboration between psychological counseling centers, libraries, and parents of dyslexic children. In the presented workshop, the possibilities of eye-tracking technology in the context of reading and dyslexia interventions will be presented. At the same time, participants will have the opportunity to try eye-tracking practically. The workshop should also facilitate a discussion based on the experience of the invited participants to identify the general needs.

The presented workshop is organized within the TAČR SIGMA project called "Reeducation of dyslexia and development of reading abilities based on eye-tracking technology".

Literature

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A Definition of Dyslexia. *Annals of Dyslexia*, *53,* 1–14. https://doi.org/10.1007/s11881-003-0001-9

Workshop: Virtual Theatre as a New Medium of Understanding Culture

Natálie Káčová, Wei-lun Lu, Alžběta Šašinková, David Košatka, Šárka Havlíčková Kysová, Čeněk Šašinka

Faculty of Arts, Masaryk University
Arna Nováka 1, 60200 Brno, Czech Republic
nkacova@mail.muni.cz, wllu@phil.muni.cz, asasinkova@mail.muni.cz, kosta@phil.muni.cz,
sarka.havlickova@phil.muni.cz, cenek.sasinka@mail.muni.cz

1 Introduction

The new wave of immersive virtual reality (iVR) headsets, the expanse in its use as well as the techno-social vision of metaverse have created a momentum for the iVR technology. iVR has gained attention for the potentials it offers to the educational as well as artistic sphere, introducing new use cases and user experiences. The paper introduces drama-based language and cultural learning mediated by iVR. The current state of art and the vision of the iVR theatre developed under a TACR research project is presented.

2 Virtual theatre, cultural and language learning: What we know so far?

Studies have analyzed the benefits of learning through drama, iVR affordances and benefits for the learning process, as well as the use of virtual environments for educational drama.

Drama in education motivates students to collaborate with each other and immerses them to new contexts and roles that may be employed to facilitate language learning (Fleming, 2006). Incorporating drama techniques to the learning process encourages students' self-expression through movement (Boudreault, 2010), reduces foreign language anxiety (Galante, 2018), enables sharing of cultural concepts by demonstrating their complexity and/or facilitating intercultural dialogue (Donelan, 2002; Heldenbrand, 2003).

A number of the attributes of iVR technology comply well with the drama in education approach and philosophy. iVR has been successfully employed for the actual production and delivery of theatre performances (e.g. Pike, 2020; Sim, 2021). Situating the user(s) to information-rich environments and simulations of a specific socio-cultural context, iVR was proven suitable for facilitating (inter)cultural learning (Shadiev et al., 2020). The strong immersiveness and interactivity of iVR further supports students' engagement and motivation (Lan, 2020; Ou Yang et al., 2020). Therefore, immersive virtual environments hold a strong potential to become safe spaces for interactive, embodied and expressive forms of cultural and language education (Lan, 2020).

So far, studies on the virtual forms of learning through drama have mainly analyzed the use of non-immersive virtual environments (e.g. Dunn, Bundy & Woodrow, 2012; Salas-Alvarado, 2021). Immersive collaborative VR has been widely researched for the specific use cases of role-playing (Atuel & Kintzle, 2021; Sapkaroski, Mundy & Dimmock, 2022) that encourages trainees to experience and rehearse social interactions, such as communicating with a patient, in a safe environment. The instructional iVR design that would conceptually merge drama in education (the artistic performance of a play) with cultural and language learning (facilitating cultural meanings and themes inside iVR while using a second language) therefore remains, to our knowledge, uncovered.

3 The vision: Acting and learning in iVR

The aim of the TACR Sigma project "Virtual Theatre as a New Medium of Understanding Culture" is to propose such a new form of cultural and language education, mediated by rehearsing and acting a play in collaborative immersive virtual reality. To achieve its goal, the project incorporates various fields of knowledge and expertise, such as VR development, HCI and UX design, research through design, scenography and screenwriting, instructional design and drama in education, as well as sinology since cultural and language learning will be designed for the Chinese cultural area.

Throughout the design process, particular attention is paid to the study of cultural realities and themes and their translation into the iVR application and the theatre scenario. We see a great potential of communicating important cultural concepts through the design of the avatars (archetypes, gestures),

environments (artefacts, surroundings), social interactions (with the other users and/or virtual agents, e.g. a virtual guide that guarantees the knowledge of cultural aspects present within the iVR application), theatre scenario (selection of the (inter)culturally relevant scenes) and the overall learning process (provoking discussions and reflections).

In order to evaluate the developed design, the course will be held where the participants (students of Chinese) will rehearse perform the theatre play in iVR together. The potential is seen in redesigning the application for different cultural/language areas and for the uses at universities and public libraries that hold a potential to mediate cultural content to the wider public.

To conclude, the artistic performance of a play in a collaborative immersive virtual environment represents an innovative and promising form of cultural and language education, combining the benefits of drama and a still emerging iVR technology for the learning process. The conceptual design of the iVR theatre introduced in this paper further elaborates these potentials, in order to immerse the users to rich cultural contexts and provide them a chance for self-expression and shared experiences in a safe virtual space.

Acknowledgements

Literature

Atuel, H. R. & Kintzle, S. (2021). Comparing the training effectiveness of virtual reality and role play among future mental health providers. *Psychological Trauma: Theory, Research, Practice, and Policy*. *13*(6), 657-664.

Boudreault, C. (2010). The benefits of using drama in the ESL/EFL Classroom". *The Internet TESL Journal, 16*(1).

Donelan, K. (2002). Embodied Practices: Ethnography and intercultural Drama in the Classroom. *Drama Australia Journal*, *26*(2).

Dunn, J., Bundy, P. & Woodrow, N. (2012). Combining Drama Pedagogy with Digital Technologies to Support the Language Learning Needs of Newly Arrived Refugee Children: A Classroom Case Study. *Research in Drama Education*, *17*(4), 477-499.

Fleming, M. (2006). Drama and language teaching: the relevance of Wittgenstein's concept of language games. *Humanising Language Teaching*, *8*(4), 97-110.

Galante, A. (2018). Drama for L2 Speaking and Language Anxiety: Evidence from Brazilian EFL Learners. *RELC Journal*, *49*(3), 273-289.

Heldenbrand. (2003). Drama techniques in English language learning. *Korea TESOL Journal*, *6*(1), 27-3.

Lan, Y. J. (2020). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology*, *24*(1), 1-15.

Ou Yang, F.-C, Lo, F.-Y. R, Chen Hsieh, J., & Wu, W.-C. V. (2020). Facilitating Communicative Ability of EFL Learners via High-Immersion Virtual Reality. *Educational Technology & Society, 23*(1), 30-49.

Pike, S. (2020). Virtually relevant: AR/VR and the theatre. *Fusion Journal, 17*.

Salas-Alvarado. J. (2021). The virtual experience of the course Drama for English as second language at the Sede del Pacífico of the Universidad de Costa Rica. *InterSedes*, *22*(46), 49-83.

Sapkaroski D., Mundy M. & Dimmock, M. (2022). Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training. *Journal of Medical Radiation Sciences*, *69*(1), 56-65.

Sim, H. (2021). Seeing Alone Yet Together: Modern-day Tiresias in VR theatre. *Performance Research*. *26*(1–2), 69-77.

Shadiev, R., Xueying W., & Huang, Y. M. (2020). Promoting Intercultural Competence in a Learning Activity Supported by Virtual Reality Technology. *International Review of Research in Open and Distributed Learning*, *21*(3), 157-174.