

Információs Társadalom

[Information Society]

A SOCIAL SCIENCE JOURNAL

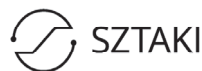
Founded in 2001

English Issue

Vol. XXIII, No. 2 (2023)

Editor-in-chief : Héder, Mihály

Managing Editor: Póra, András



Published by INFONIA Foundation

Principal sponsor: Budapest University of Technology and Economics, Faculty of Economic and Social Sciences

Technical partner: SZTAKI

Editorial Board:

Székely, Iván (Chair)

Alföldi, István
Berényi, Gábor
Bertini, Patrizia
Bethlendi, András
Csótó, Mihály
Demeter, Tamás

Molnár, Szilárd
Pintér, Róbert
Rab, Árpád
Z. Karvalics, László

Copy Editor: Tamaskó, Dávid

ISSN 1587-8694 (Print)

ISSN 2063-4552 (Online)

ISSN-L 1587-8694

Produced by Server Line Print & Design, Budapest

The journal Information Society (In Hungarian: Információs Társadalom, abbreviated as InfTars) aims to provide a platform for research and discussion of the theories and applications of Information Society Studies. Currently every second issue is published in English, the rest are in Hungarian.

The journal is fully Open Access and freely available at <https://infmars.infonia.hu/>

InfTars is indexed in both the *Web of Science Social Sciences Citation Index and Scopus*, and all titles are automatically submitted to *Crossref*.

Since Vol. VIII, No. 1 (2008) the journal has been included in the Thomson Reuters index (Social Sciences Citation Index, Social SciSearch, Journal Citation Reports/Social/Sciences Edition)

E-mail: infmars-lapman@ponens.org

CONTENTS

LECTORI SALUTEM

7

**DALMA LILLA DOMINEK, NÓRA BARNUCZ, ERNA URICSKA,
LÁSZLÓ CHRISTIÁN**

Experiences of digital education from the students’ perspective

9

According to some literature about communication in a target language, digital competency and creativity are highly important, and achieving this is exceedingly feasible with the methods employed in experiential education. The aim of the present study was to identify the advantages and disadvantages of the online and offline digital education of students at the University of Public Service (hereinafter: UPS) in Budapest, Hungary, utilizing planned experience-based research. The relevance of the research covers two independent and separate quantitative research themes: 1. digital education during the first period of the coronavirus pandemic; and 2. online quantitative exploratory research. The results of the research support the planned experience-based research and confirm it is suitable for measuring the “flow state” of students, where flow is the optimal state where someone is fully engaged in a task at hand with little distraction. It was revealed that students can reach a high flow state when participating in a digital learning environment, which facilitates high motivation to participate in the lessons and has positive effects on the development of the students’ communication and digital competences.

**ZOLTÁN GELLÉR, DRAGANA MILIN, TANJA JEVREMOV,
DUBRAVKA VALIĆ NEDELJKOVIĆ, MIRJANA IVANOVIĆ**

Non-ICT Students’ Familiarity with Basic Computer Concepts at the Faculty of Philosophy at the University of Novi Sad

25

As a continuation of our previous paper, the aim of this research was to determine the level of familiarity of the freshmen at the Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, with the basic concepts related to computers. The sample included the same 568 respondents as in our first study, and the same two control groups. Familiarity was analysed with 14 questions from the field of computer literacy, both at the level of individual questions and in relation to the total number of correct answers. The results were also analysed in relation to the condition of passing the test from the appropriate European

Computer Driving Licence module. No student answered all the questions correctly, which clearly indicates the need to educate students on basic concepts related to computers. The obtained results can be used to better adapt teaching to the students' needs and their prior knowledge.

DALMA GESZTEN, BALÁZS PÉTER HÁMORNIK

Can team communication obscure severe usability problems?

48

Team communication is a significant topic of study in the psychological literature, but it is a less researched area in relation to software usability. In this article, as part of a larger research project, we investigated team communication to explore the relationship between usability problems and team communication patterns. The research question was examined in two laboratory experiments as part of a collaborative software evaluation process. For sequential analysis of team communication transcripts, we created a custom code system based on previous literature. The results confirm that teams that experience special types of usability problem show different communicational patterns from teams that experience no such problems. The results also show high reliability of the new code system. Further research is needed to explore the relationship between usability problems and team communication patterns in different settings.

JERNEJ KALUŽA

Hume's Empiricism versus Kant's Critical Philosophy (in the Times of Artificial Intelligence and the Attention Economy)

67

The article exposes how one of the most fundamental oppositions in the history of modern philosophy – the opposition between Hume's empiricism and Kant's critical philosophy – is embedded in the current debate on the impact of artificial intelligence (in particular, the algorithmic selection of content) on human society. Hume's empiricism – with its deduction of subjectivity based on a process of habituation – corresponds to the functioning of recommending algorithms, while Kant's idea of autonomous subjectivity corresponds to the ideals underlying today's ethical attempts towards the regulation of artificial intelligence. According to such ethics, the use of empirical data can endanger humans; whereby our attention can be easily caught by sensationalist content and our autonomy replaced by the agency of machinic intelligence. However, as argued in the present article, such ethical positioning also reproduces the gap between the empirical reality and normative principles, which is why transcendental (Kantian) ethics should be supplemented with Hume's immanent and practical reasoning.

BEÁTA LAKI

Self-protective versus utilitarian autonomous vehicles

83

The following examination focuses on the moral dilemmas surrounding the comparison of self-protective and utilitarian autonomous vehicles. These vehicles can be programmed to prioritize the safety of passengers in an accident or prioritize the greater good to save more lives. The essay will explore various ethical questions such as evaluating the numbers game approach, analyzing the principles of beneficence correlated with social inequality, and interpreting the principle of autonomy in the context of autonomous vehicles. Additionally, the examination will consider the harm–benefit ratio. In the sense of methods, this analysis uses the classical issues in a new way and provides recommendations for decision-makers to consider.

JUHO LINDHOLM

Technics as Hermeneutics

95

In this paper, I outline, how technics can be understood as a special case of hermeneutics. My argument hinges on the concept of practice, which, according to Joseph Rouse, is hermeneutic activity. I show how Rouse’s practical hermeneutics and Don Ihde’s material hermeneutics can be considered to be founded on the pragmatist theory of meaning. For John Dewey, interpretation is thoroughly technical.

PATRIK REIZINGER

The Falsificationist View of Machine Learning

117

Machine learning pushes the frontiers of algorithmic achievements, though the striving for state-of-the-art performance often obscures the fragility of enforcing decisions amid uncertainty. This paper interprets machine learning within Karl Popper’s epistemology. We assess machine learning paradigms’ fit for falsificationism and argue that the new interpretation can improve robustness. Though the price is to accept unambiguous decisions, the restriction of the hypothesis space still adds value. The context for our work is established by comparison with similar techniques and highlighting its limitations.

GIACOMO FIGÀ TALAMANCA

Frameless Folk Psychology

128

In this paper, I will argue that the rise in hostility and polarization on social media is explainable by taking into account a radical difference between online and face-to-face interaction. In everyday offline environments, socially shared

and context-dependent norms frame the understanding of other people's minds based on their behavior. I will argue that, on social media platforms, social cognition is distorted thanks to two deliberate design choices that are a means for financial gain for the platform's designers: namely, the lack of socially shared norms on these platforms (entailed by what is known as context collapse) and their interfaces' extreme user-centeredness. I will argue that such design features not only cause frustration in the understanding of others but encourage testimonial injustice in interaction.

ERNA URICSKA

Trust is in the (linguistic) details: policing digilect as a communication practice

146

The use of social network sites by law enforcement agencies takes place between the police as a service provider and the members of the public. Site administrators need to find a communication style that is understandable to the broadest audience. In the course of my research, I examined the entries created on the Instagram profiles of the Budapest Metropolitan Police Headquarters (Budapesti Rendőrfőkapitányság, hereinafter: BRFK) and the Hungarian Police (Magyar Rendőrség, hereinafter: MR) in two sampling periods (sampling took place in the first month of both profiles, and a year later) using the methodology of content and discourse analysis. Comparison of the results has shown that there is a more casual style of gaining the trust of the population. Appropriate and regular partnership communication with civilians can be used effectively in crime prevention and detection, and can also have positive effects on security.

LECTORI SALUTEM

The editorial board welcomes the readers of the No. 2 issue of 2023!

In the first paper, Dominek et al. present their research about digital education from the students' perspective. Using planned experience-based research, they tried to identify the advantages and disadvantages of online and offline digital education of students at the University of Public Service in Budapest, Hungary. According to the main findings, the method is suitable for measuring the "flow state" of students – which facilitates high motivation to participate in the lessons and positively affects the development of the students' communication and digital competencies.

In the following study, Gellér et al. measured the non-ICT students' familiarity with basic computer concepts at the Faculty of Philosophy at the University of Novi Sad, Serbia, by a survey-based analysis. The aim was to determine the level of familiarity of the freshmen (568 respondents, 14 questions) with the basic concepts related to computers as a continuation of their previous research. No student answered all the questions correctly, which indicates the need to educate students on the issue.

The third paper (Geszten and Hámornik) investigated team communication to explore the relationship between usability problems and team communication patterns with two laboratory experiments as part of a collaborative software evaluation process. The results confirm that teams that experience particular types of usability problems show different communicational patterns from teams that experience no such problems.

Turning to the more philosophical issues, Kaluža's article exposes how one of the most fundamental oppositions in the history of modern philosophy – the opposition between Hume's empiricism and Kant's critical philosophy – is embedded in the current debate on the impact of artificial intelligence (in particular, the algorithmic selection of content) on human society. He argues that this embedded opposition reproduces the gap between empirical reality and normative principles, which is why transcendental (Kantian) ethics should be supplemented with Hume's immanent and practical reasoning.

In the fifth study, Laki immersed into the moral dilemmas surrounding the comparison of self-protective (saving the passengers) and utilitarian (prioritizing saving more lives) autonomous vehicles. The essay explores various ethical questions, such as evaluating the numbers game approach, analyzing the principles of beneficence correlated with social inequality, and interpreting the principle of autonomy in the context of autonomous vehicles – also considering a harm-benefit ratio providing recommendations for decision-makers.

Staying on the philosophical track, Lindholm's research outlines how technics can be understood as a special case of hermeneutics. The argument shows how Rouse's practical hermeneutics and Don Ihde's material hermeneutics can be considered founded on the pragmatist theory of meaning (also that for John Dewey, interpretation is thoroughly technical).

The seventh article (Reizinger) is about the falsificationist view of machine learning. The paper interprets machine learning within Karl Popper's epistemology, as

sessing its fit for falsificationism, and also states that the new interpretation can improve robustness. Though the price is to accept unambiguous decisions, the restriction of the hypothesis space still adds value.

The penultimate paper from Figà Talamanca claims that the rise in hostility and polarization on social media is explainable by two deliberate design choices for the platform owners' financial gains: the lack of socially shared norms and their interfaces' extreme user-centeredness. According to the author, these features not only cause frustration in understanding others but encourage testimonial injustice in interaction.

Finally, Uricska's research is about gaining public trust in the communication of the police. The author examined Instagram entries of the Budapest Metropolitan Police Headquarters and the Hungarian Police in two sampling periods using content and discourse analysis methodology. The results: there is a more casual style of gaining the population's trust. Appropriate and regular partnership communication with civilians can be used effectively in crime prevention and detection and can positively affect security.

The editorial board wishes you a pleasant time reading this issue.

Experiences of digital education from the students' perspective

According to some literature about communication in a target language, digital competency and creativity are highly important, and achieving this is exceedingly feasible with the methods employed in experiential education. The aim of the present study was to identify the advantages and disadvantages of the online and offline digital education of students at the University of Public Service (hereinafter: UPS) in Budapest, Hungary, utilizing planned experience-based research. The relevance of the research covers two independent and separate quantitative research themes: 1. digital education during the first period of the coronavirus pandemic; and 2. online quantitative exploratory research. The results of the research support the planned experience-based research and confirm it is suitable for measuring the “flow state” of students, where flow is the optimal state where someone is fully engaged in a task at hand with little distraction. It was revealed that students can reach a high flow state when participating in a digital learning environment, which facilitates high motivation to participate in the lessons and has positive effects on the development of the students' communication and digital competences.

Keywords: *digital education, flow, multimedia, experiential education, social network sites, higher education*

Author Information

Dalma Lilla Dominek, University of Public Service, Faculty of Public Governance and International Studies, Department of Social Communication, Assistant Professor
<https://orcid.org/0000-0002-8460-640X>

Nóra Barnucz, University of Public Service, Faculty of Law Enforcement, Department of Foreign Languages for Specific Purposes, Assistant Lecturer
<https://orcid.org/0000-0001-7832-3465>

Erna Uricska, Corvinus University of Budapest, Doctoral School of Sociology and Communication Science <https://orcid.org/0000-0002-5371-0650>

László Christián, University of Public Service, Faculty of Law Enforcement, Department of Private Security and Municipal Policing, Associate Professor, Head of the Department, Vice Rector <https://orcid.org/0000-0001-9809-4890>

How to cite this article:

Dominek, Dalma Lilla, Nóra Barnucz, Erna Uricska, László Christián. “Experiences of digital education from the students' perspective”.

Információs Társadalom XXIII, no. 2 (2023): 9–24.

==== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.1> ====

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

One of the most important tasks of higher education is to develop students' competences that can be used in the labour market, particularly students' active language abilities. Such language skills development is essential at all national and international institutions of higher education. In order to develop these skills, teachers commonly aim to use the flow channel in education (Csíkszentmihályi 2010; Dominek 2021a). Here, "flow", such as in the saying "in the flow" (asking to someone being "in the zone"), is a mental state in which a person is totally absorbed in the task at hand, and where that person has an energised focus and full involvement in the task to the point where they can lose track of both time and self. In the teaching context, when students do not perceive spending time in the lesson or focus on the environment surrounding them in the class, they can get into the flow channel, and then teachers have a great opportunity to develop their creativity by grabbing their undivided attention and engagement (Dominek 2021b).

The purpose of this study was to identify the advantages and disadvantages of online and offline digital education according to the students' perceptions as part of a planned experience-based research. Its theoretical background and the role of multimedia and mobile devices in education have already been published (Barnucz and Uricska 2020, 2021). In this paper, first we present the theoretical background, giving the definition of the concept of flow and its relevance in education, and second, we present the qualitative and quantitative results of our study into the effectiveness of digital education delivered online and/or offline.

The theoretical background and research introducing the experiences of digital education from the perspective of students largely confirmed the relevance of the planned research. In the near future, we aim to carry out classroom research by creating a digital and multimedia-based environment with the application of augmented reality (AR) in line with prior literature (Czékman 2017a; Aknai, Czékman, and Fehér 2021) and considering the HY-DE model (an acronym derived from *HY*per and *DE*eper attention, as outlined in Dani 2014) in the classroom. The research cohort comprised students at the Faculty of Law Enforcement, UPS (hereinafter: FLE) who were studying English for law enforcement. The purpose of the research is to assess and ultimately improve the level of the students' digital and verbal communication and vocabulary knowledge in the framework of English for Specific Purposes (ESP) classes. In addition, the pilot research also focuses on the development of the flow state in students.

2. Multimedia – Mobile Learning – Flow

The language learning methods mentioned above can be applied in *multimedia environments*, which are highly prevalent in today's education system. In terms of the effect of multimedia on learning foreign languages, Mayer (2001), based on Wittrock's

generative learning theory and Paivio's dual coding theory (Paivio 1969; Paivio and Desrochers 1980), stated that memory and cognitive abilities have an effect on sensory impression. According to such theories, when students use both verbal and visual systems, they can code the information they are given, and the information that they have gained is then learned and stored in their long-term memory. In contrast, when students use only the verbal system for learning, it is less likely that the gained information will be stored in their long-term memory. A curriculum built on multimedia can decrease the pressure on the memory, and this has been found to be particularly important in the acquisition of a foreign language (Zhongyuan 2013; Teng and Zhang 2021).

Not only multimedia, but *mobile devices*, such as tablets and/or smartphones, are also effective tools in this system (Czékman 2017b). National and international researchers agree that "mobile learning" means those learning materials and experiences that are available anywhere and at any time, but it is important to remember that "*mobile learning is an activity controlled not by technology but by students*" (Turner 2012). Thanks to the use of mobile devices, there is a possibility to create new mobile learning environments (MLEs) that allow a focus on, among others, thinking processes at an advanced level, the development of 21st century skills, personalized learning environments, and the establishment and maintenance of motivation (McQuiggan et al. 2015), as well as for taking advantage of compensating effects for overcoming social disadvantages.

Just as these methods described above have been applied in the field of environmental education, similar positive results could be achieved in ESP classes for law enforcement. The target categories in environmental education at the individual level that have been studied in the literature include an openness to facing problems (i.e. the use of metacognitive models during language learning by Chamot, Uhl, and El-Dinary 1999; Barnucz and Uricska 2021), and developing appropriate knowledge and skills, and a proper attitude, and active participation (motivation), as well as the sustainability of the whole production and consumption system at the level of societies (Dominek 2021c). In this context, the development of target categories in environmental education can also be realized in a multimedia environment and the experience of "flow" can be expected to energise the learning process. It should be mentioned that flow is not directly proportional to happiness, and we look to positive psychology that deals with the science of happiness for a greater understanding of this. In positive psychology, the "PERMA" model, in which there are five different factors (1. Positive feelings, 2. Engagement, 3. Relationships, 4. Meaning, and 5. Achievements) that are building blocks that contribute to a person reaching a state of wellbeing and happiness (Dominek 2021d). Flow is a part of this psychological research area. When students experience a state of flow, they exclude the environment and time factors around them. During their activities when they are in the flow, the students can carry out intense work, which means they can address challenges and find a balance, while they are able to enhance their skills through their increased focus, including their motivation, patience, creativity, problem-solving, independent thinking, to reach optimal peak performance.

The concept of flow derives from the theory of creativity (Dominek 2021d). According to Rhodes' (1961) theory, there are four aspects that influence creativity, often called the four Ps: the person, the process, the press (i.e. the external environmental effects), and the product (i.e. the end results). Creativity can be developed, and this is the essential point where the theory of Csíkszentmihályi's flow is derived from (Dominek 2017). In his work "Flow", the author explained that when people experience a high degree of joy, their concentration can better focus on a challenging task, and they can achieve the perfect experience at this point. Csíkszentmihályi's research (2010, 2015) also addressed how ordinary people experience flow mostly during their work. Work can become a real creative activity by providing the right experience for a person. New feelings and desires can be formed by ordinary people through their experiences, while immersion in variety, the exploration of novelty, and taking part in the decision-making process can be experienced by people through work. Such experiences can lead to great happiness in people. Csíkszentmihályi (2015) wrote how it is possible to reach an optimal experience when limits and barriers disappear, and people can tap into a new dimension. Boswijk and co-authors (2007) wrote about the richness of experiences and stated that the most important thing in an individual's life is to gain experience. According to Csíkszentmihályi, people can exceed the things expected of them when they are in the flow channel and they can gain such experiences that they may not have expected before (Dominek 2020).

As opposed to this, some studies prove (e.g. Brown 1994; Kelemen and Talabér 2014) that the language learning process is not considered an ideal experience for language learners in several cases. It has also been proven that there exist breakpoints in the development of communication in a target language, which can have both psychological and pedagogical reasons. According to national and international research (e.g. Dicheva et al. 2015; Niyetbaeva et al. 2016; Borszéli 2019; Molnár and Uricska 2019; Barnucz and Fónai 2020), while there is a wide variety of pedagogical reasons for such breakpoints (i.e. the lack of appropriate methodology of the teachers, low institutional infrastructure, etc.), the psychological reasons for breakpoints can be derived from a lack of the development of creativity (Dominek 2021a). In our earlier research into students' level of creativity in higher education, with a cohort of 3509 students, we found that the strongest components of the students' creativity were playfulness and humour, while the weakest characteristics were nonconformity and impatience (Dominek and Ceglédi 2021).

Overall, according to the literature, we can say that creative-based pedagogy is indeed able to get the participants in education into the flow channel. Consequently, the pedagogical model of flow can be an effective method for promoting creative-based learning. Given this, Dominek's new educational model, the flow-based pedagogical model (Dominek 2022), which includes and involves creativity, flexible thinking, playfulness, humour, and brainstorming, would be a valuable tool to integrate into higher education to improve students' learning. Her empirical results show that when teachers use these elements, the students can get into the flow channel (Dominek 2020, 2021c).

3. Introduction to the Research: Methods and Sample Cohort

The relevance of the present research is given by two independent and separate quantitative research studies involving the students of UPS. The aim of this study was to prove the relevance of the research built as part of the planning of the curriculum for the development of vocabulary and digital competence in a student-centred learning environment (see in the introduction of the planned research) (e.g. Wurdinger and Carlson 2009; Wurdinger and Allison 2017). We would like to contribute to the strategic objectives of the Creative Learning Program of UPS, including the realization of a pedagogical, methodological change in education concentrated on the effective development and evaluation of students' abilities, the mentoring of the individuals' ways of learning, and the maintenance of the professional communities built on the individuals' professional knowledge. We intend to join the direction of the pedagogical development along with the objectives we mentioned above and the three principles of the Institutional Development Plan (2020–2025) of UPS, namely *community learning – creation – individual development*.

- One of the research studies (hereinafter: research into digital education) observed the experience of digital education among the students during online teaching. An online questionnaire was sent to all the students of UPS (N=5670). The leader of the research was Dalma Lilla Dominek. The individual research took place at UPS between August and September 2021.
- The other pillar of the relevance of the present research was an online questionnaire that included two surveys; one of them observed the role of social networking sites played in language use and in students' obtaining information and doing distance learning during the pandemic period, and the other observed the usefulness and effectiveness of education supported by Internet platforms among the students studying English for law enforcement (N=180).

The research into digital education – based on an online questionnaire dealing with the students' experiences of online teaching and learning – was carried out at the end of the academic year 2020–2021, and at the beginning of the academic year 2021–2022. The questionnaire covered three dimensions. The database contained data from 702 respondents. The first group of the questions measured creativity (Dominek 2021a), the second focused on the students' flow experience (Dominek 2021d), and the third one studied the students' experience of digital education, and the results are presented in the study. During the research, we planned a representative sampling procedure for the students, and thus the results can be considered representative. In addition to other researchers (e.g. Ürmösné and Kovács 2022; Szabó and Buda 2022), we examined what the students think about digital education and how they could acquire the learning material they needed during the pandemic period. Moreover, we asked how they could imagine education in the future. According to our hypothesis, the usefulness of digital education during the pandemic was unquestionable, however, effective knowledge transfer from the instructors requires further methodological development (Sung, Kuo-En, and Tsu-Chien 2016). We also assumed that the students would see the usefulness of digital education mostly

in a flexible and liveable organization of everyday life (Mathes 2019; Soffer, Kahan, and Nachmias 2019).

The questionnaire on the utilization of social network sites in the teaching process covered five dimensions and consisted of 16 questions. We conducted non-full-scale research. The research was carried out in the spring and autumn semesters of the academic year 2020–2021. The self-administered questionnaire was completed by 67 persons. The validity of the research was not influenced by the digital education during the pandemic as various online sources and programmes, digital platforms, and social network sites had also been used in the lessons before the first wave of the pandemic.

The questionnaire for the research into the use of digital platforms covered five dimensions and consisted of 21 questions. The self-administered questionnaire was completed by 68 students. When compiling the questionnaire, we identified questions related to the background information of the respondents, and the habits of the use of the Internet and e-materials. In the research, e-materials were considered as learning materials made on or downloaded from different Internet platforms by the teacher (e.g. flashcards to develop vocabulary). It is a fact that because of the frequent use of Internet-based platforms (e.g. Kahoot, Quizlet, Mentimeter, Padlet) and social network sites (e.g. Instagram, Facebook) in ESP classes, the switch to digital education was more flexible during the pandemic. In this case, we did not plan a representative sampling procedure, and thus the results are not considered to be generalized; however, they are suitable for exploratory research.

3.1. Descriptive statistics of the research

In the research into digital education, the distribution of students by gender was relatively well balanced, whereby 44.3% of the respondents were female, while 55.7% were male. The distribution by age showed that the questionnaire was completed mostly by full-time students, with 42.3% of them between 18–25 years old and 15.8% of them between 26–33 years old. Also, 41.9% of the students attended correspondence training; 7% of them attended courses with any specialization, and 9% of them participated in any training course.

In the case of the other two surveys, the distribution of the students by gender was also relatively well balanced. Regarding the research into the role of social network sites, 46.3% of the respondents were female, while 53.7% were male. Regarding the research into the use of Internet platforms, 44.1% of the students were female, while 55.9% were male.

The distribution by age shows that the questionnaires were completed mostly by full-time students. In the case of the research into the role of social network sites, 83.58% of the respondents were between 18–23 years old, while 11.94% of them took part in correspondence training, and the distribution of their age ranged between 24–34 years old. Also, 4.48% were between the ages of from 35 to 48 years old. The average age of the participants was 21.53 years old. In the case of the research into the use of Internet platforms, the rate of full-time students was 82.4% who were

aged between 18–23 years old, while only 8.8% of the respondents between the ages of 24–34 years old attended correspondence training. The average age of the participants was 22.63 years old.

4. Qualitative Results

In this section, we focus on the students' experiences of digital education during the pandemic based on the open questions of the third block in the questionnaire (see chapter 2). We analyse the manifest contents of the answers with content analysis.

One of the most important results was that 41.7% of the respondents would like hybrid education in the future, that is, a combination of traditional and digital education: *"I prefer digital education in a hybrid way, but in my opinion, face-to-face consultations are also necessary."* *"I would gladly take part in hybrid education in the future in a correspondence training. It is easier to complete when a person has a job and a family [...]."* Only 27.4% of the respondents preferred traditional education, while others (30.9%) imagined their future university education in a digital way.

Overall, 67.4% of the respondents declared that the online teaching and learning during the pandemic had been appropriate at the university, and 63.4% of them would like to continue their studies in a digital way. The following comments verify the satisfaction of the students: *"As it is a correspondence program, and with the introduction of digital education, the transfer of the educational curriculum has become much more effective and more successful, regarding time management. It is accomplished through multiple channels, much more information is available, and the lectures can be watched again, which is impossible in case of traditional education. I am less tired, because I have to travel from the countryside so I can avoid several hours of travelling. In the winter period, I was often late because of the break in the overhead line, freeze or a mechanical failure. There is no chance for that in digital education. It costs less money, because I do not have to pay for fees and accommodation."* *"In certain programs there is no need to be present on lessons, the same knowledge can be given online."* *"Since the introduction of digital education my life has been much more balanced. Now, before the morning lectures I can run or do exercises for an hour on a daily basis."*

The next closed question concerned the quality of the digital lectures and practical courses, and included an open question, which gave the student an opportunity to explain their answer. Overall, 88.6% of the students were satisfied with the digital availability of the lectures, and according to their feedback, they were able to listen to the uploaded lectures online with no difficulties. Based on their arguments, it can be stated that the education was delivered at a high level in the online environment, while the students did not miss lectures, and all the technical conditions well supported the teaching and learning.

In contrast, 55.4% of the respondents were not satisfied with the digital practical courses. According to their feedback: *"Practical lessons of my specialization can rarely be called practical, I would rather call them lectures."* *"In practical lessons, practice and not theory should be in focus."* *"Almost none of my practical lessons have been kept online properly, in my opinion, they should have been kept face to face, if it is possible."*

“Practice cannot be transferred digitally. The main point of practice is that students can do the steps themselves, use machines, devices and so on. It can be presented online like a theory, and it could not be transferred into practice later almost impossible.”

The students were asked to present examples of an appropriate and an inappropriate practice. The students mostly emphasized the importance of ESP classes, from which many examples could be given relevant to our planned research. As one example, some students mentioned that they had been taking part in ESP lessons with the camera and sound turned on, or they had been working in Team-rooms or how it had been good to solve Quizlet tasks. Not only ESP classes but also practical lessons for practising techniques for negotiations were also mentioned. The students worked in pairs in classes, and they had to organize a conference, for example. They also emphasized that it had not been appropriate when the teacher had been talking continuously for 90 minutes without involving the students. In contrast, when students were given tasks that they had to solve together, they were motivated to do the task and to participate in class. Their expectations had been identified in line with the teachers, and thus the questions of flexibility, the expanding of creativity, the use of appropriate and effective communication, and more shared content, videos, interactive tasks (e.g. voting, quizzes in classes) were also important for the students.

The students also expressed their views about their expectations and critical comments about teachers. In their opinion, it is important that *“teachers must be trained to handle online educational programmes properly and professionally. Many of the teachers are still not capable to run the given programmes properly.”* Besides, *“They must be student-centric, they must listen to students’ needs and organize lessons by students’ best interests.”* It is important that teachers *“should be creative, they should not only explain the curriculum, but give practical examples, show presentations to demonstrate the curriculum. They should inform the students about the requirements of tests, assignments and exams in writing, as soon as possible.”* The importance of written notifications was also emphasized by one student: *“Written notification should be compulsory outside of class as well, as some of my peers are unable to attend the class and miss important information that the teacher is given verbally without any written information before and after the class.”*

Last, but not least we asked how suitable digital education was considered according to their experiences. Overall, 70.7% of them were satisfied with it, while one of them said that it was like the *“third world”*, or another one said that *“digital education is a kind of convenience in a broad sense. Lessons for correspondence students last from Friday afternoon to Friday evening and all day on Saturday. It is twice or three times more difficult to be there in the lesson than in case of digital education. Those, who have kids, nurse their parents, live on the other side of the country, are sick at home, have financial problems and cannot pay the fees of travelling and accommodation, they are all grateful for digital education.”*

Moreover, the most successful initiative is summarised in the following paragraph: *“recording and publishing lectures on webinar (can be watched any time); make Probono materials public (outstanding curriculum); the introduction of compulsory use of Moodle both for the instructors of the departments and students (all requirements, curriculums, teams links can be found in one place).”*

And finally, some proposals were also formulated by the students: *“For me, digital education means the 21st century. I do not say that traditional, face-to-face education is out of date, but its extent should be reconsidered and give more space to digital education in correspondence training. If it is possible, please do not take these from students who take part in distance education, but keep them informed. I would also support that the traditional lectures were recorded and they would be available online linked to the Neptun code and the possession of a medical certificate for those students who were absent (sickness, maternity leave, other obstructions).*”

5. Quantitative Results

In this section, the role of social network sites and Internet platforms in language teaching is examined. This section contains the basic statistical results of the two exploratory research studies above, which sufficiently contribute to the relevance of the planned research. A more specific, deeper statistical analysis will be carried out in another study.

The relevance of the research plan (see in the next chapter) is presenting the students' positive feedback in connection with the digital platforms and social network sites used within the framework of ESP classes. In the case of the habits of obtaining information online, 63% of the respondents chose social network sites, 33% read online news sites, and only 4% did not read online sources. In the research into using social network sites in class, 94% of the respondents read some kind of social network sites to obtain information: more than 90% reported using Facebook, 75% YouTube, and approximately 70% Instagram. Social network sites have primary importance for students in obtaining information (Uricska 2022; Uricska and Molnár 2022). The third unit of the survey examined the role of social network sites from the viewpoint of language learning due to the coronavirus, and the new circumstances it prompted. The results show that the use of social network sites resulted in vocabulary expansion both in English and Hungarian. This result was also justified by the vocabulary collection written during the pandemic period (Uricska 2021; Barnucz 2021, 2022).

In connection with the research into using Internet platforms in class, the frequency of Internet use was examined on a five-grade Likert scale (from 1 = not at all to 5 = almost every day). We created a dummy variable in order to make the interpretation easier (0 = never used; 1 = often used). Overall, 97.1% of the students claimed that they often use various online sources in ESP classes. After that, we were interested in what purposes the online sources/curriculums had been used for in the ESP classes. The results showed that they were used the most for practising (91.2%), for vocabulary development (69.1%), for individual learning (63.2%), and for solving online tests and making presentations (54.4%). On the contrary, online sources were the least used for project work (42.6%), tests (39.7%), entertainment (29.4%), and individual project work (19.1%) (note, results all below 50%). The frequency of the listed activities refers to the diverse methods applied in ESP classes and by ESP teachers with many methodological instruments available.

6. Discussion

The aim of this study was to prove the relevance of the classroom research carried out in a digital learning environment with two basic research studies: (1) research into the students' experiences about the digital education they received during the pandemic; (2) research with an online questionnaire, including research focusing on the roles of social network sites in education, obtaining information, and learning habits during the pandemic period and research regarding the usefulness of education supported by Internet platforms. This study presented mostly the students' experiences concerning the digital education they received during the pandemic and consisted of basic statistics of using social networks sites and Internet platforms in education.

According to our first hypothesis, the usefulness of digital education during the pandemic is unquestionable; however, effective knowledge transfer requires further methodological development. Both parts of the first assumption were confirmed, as the majority of the students (60%) thought that the usefulness of digital education is indisputable from several aspects. The students emphasized the easier predictability, better scheduling, family reasons (parenting, nursing elderly parents), lower costs, etc. as particular benefits of digital education. A minority (approximately 30%) would prefer hybrid education.

The second part of the hypothesis was also justified by the result that nearly 90% of the students were satisfied with the digital availability of lectures. Others (55% of the students) pointed out that currently teachers do not have a satisfying toolkit to operate digital education effectively, and because of this, they expressed some dissatisfaction regarding the digital practical classes. The respondents also provided some critical comments, expectations, and suggestions concerning mostly the importance of more flexible teachers' duties, giving information about dates, and the evaluation of tests and assignments.

According to our second hypothesis, the students considered the usefulness of digital education mostly to be in the availability of a more flexible organization of their daily routine. This viewpoint seemed to be verified, as the students reported they were comfortable with the digital education despite the initial difficulties that might have been derived from their inexperience with the unfamiliar situation, and the lack of methodology in place initially.

The descriptive results of the other two research studies presented in the paper (the use of social network sites for obtaining information, the use of Internet platforms in class) emphasize that the key to the development of a country is the quality of the knowledge of its citizens (Hanushek 2019) and to achieve this, the development of 21st century competences in class is essential both in online and offline education. The expectations of the students also verified the demand for developing creativity, innovation, critical-thinking, and problem-solving on the one hand, and for the development of communication and solution-centred cooperation on the other hand (Barnucz and Uricska 2021).

The results show that two essential circumstances must be seen to reach an authentic result in the process of teaching/learning. First, the learning habits of gen-

eration Z (i.e. those born between 1995 and 2000) that have already appeared both in public and higher education have changed compared to previous patterns, and teachers in all forms of education need to respond to this. Second, one of the most important tasks of higher education is to develop the competences of students so that they can use these skills well in the labour market. In fact, establishing the congruence of higher education and the labour market deserves particular attention (e.g. Bocsi 2013; Christián and Erdős 2020; Christián 2022; Christián, Hautzinger, and Kovács 2021), because new positions have appeared that need digital skills (e.g. digital police communications officers) (Metropolitan Police Service 2020). The development of digital literacy and communication is essential (Uricska 2020a, 2020b) to carry out these duties. Although our planned research requires offline education, it contributes to the expansion of the teachers' methodological toolkit. In addition, the research contributes to the suggestions of the methodological development of the ESP language teaching, and can be a good starting point for the creation of a more student-centred learning environment, where students will expectedly feel well in themselves to be able to meet challenges that encourage them to think, solve problems, and cooperate.

7. Future Perspectives

In the near future, we aim to assess the digital competence and vocabulary of students studying English for law enforcement, where the focus will be on the digital/multimedia environment created by applying the AR and HY-DE model in classes. We will try to identify the pedagogical reasons for the breakpoints in ESP teaching and learning with the research, and make suggestions on how to eliminate them. During the research, we will test the tools so that the relevant breakpoints of teaching/learning foreign languages and the development of communication can be recognized and – within the system of each method – we will make suggestions on how to eliminate difficulties (e.g. difficulties with the development in acquiring vocabulary).

In this research, we examined the judgement of the utilisation of teaching English for law enforcement supported by digital devices among full-time students (N=100). Pilot research was carried out in the 2021/2022 academic year at the Faculty of Law Enforcement, but it is an on-going research, and we are planning to expand the research to three other faculties of the university in the following two academic years. The research consists of three steps: (1) classroom research; (2) quantitative research (questionnaire); (3) qualitative (interviews in focus groups, in-depth interviews). The measurement tools are: written pre- and post-tests to test the vocabulary knowledge of the students and a flow-test that is filled out at the end of the test lessons to measure the flow state of the students both in the test and control groups. The measurement is based on a test for experiencing flow that was elaborated by Mihály Csíkszentmihályi, and it has been applied just a few times for measuring the degree of such experience (Szabó 2014; Dominek 2020) because it is difficult to find a more obvious, adequate, and trustworthy procedure for measuring flow.

In our research, we try to answer the question of how the application of AR and the HY-DE model in classes can affect the development of the students' digital competence, communication in a target language, vocabulary knowledge, and the activation of their motivation in class. We also examine whether students can experience flow with a positive psychology in the digital environment, or whether they may be able to reach a higher performance in the flow state. We research what areas of language skills can be developed by digital technology, and how the quick development of information technology can affect language teaching, and the acquisition of ESP. We assume that the digital and verbal communication of the students can be developed due to the involved language learning methods, and that the application of digital technology in classes can affect the motivation of the students and the development of the language skills in a positive way (Barnucz 2019a, 2019b; Uricska 2020c; Uricska and Suták 2022). We reveal that the students who are treated during the research can experience higher flow than those in the control group (Dominek 2020, 2021a).

The responses of the students may reveal a more intensive demand for developing a situated and constructive learning perception and at the same time how to reach a state of flow (Nahalka 1997; Sharma and Gupta 2016).

References

- Aknai, Dóra Orsolya, Balázs Czékmán, and Péter Fehér. "Kiterjesztett valóság (AR) alkalmazások használata és készítése az iskolában." In *Iskola a társadalmi térben és időben VII.*, edited by Patrícia Maisch, Nikolett Márhoffer, Zsófia Molnár-Kovács, Nikoletta Szekeres, and Karolina Szűcs-Rusznak, 150–156. Magyarország, Pécs: PTE BTK „Oktatás és Társadalom” Neveléstudományi Doktori Iskola, 2021.
- Barnucz, Nóra. "IKT eszközök szerepe az angol nyelv oktatásában." *Educatio* 28, no. 2 (2019a): 403–414.
- Barnucz, Nóra. "IKT-eszközökkel Támogatott (Rendészeti) Nyelvoktatás." *Magyar Rendészet* 19, no. 4 (2019b): 15–31.
<https://doi.org/10.32577/mr.2019.4.1>
- Barnucz, Nóra. "A kiterjesztett valóság alkalmazása a nyelvoktatásban különös tekintett a rendészeti szaknyelvre." *Educatio* 29, no. 4 (2020): 644–652.
<https://doi.org/10.1556/2063.29.2020.4.9>
- Barnucz, Nóra, and Mihály Fónai. "Az IKT-eszközök használata az idegennyelv oktatásban hátrányos helyzetű tanulók iskoláiban." *Információs Társadalom* 20, no. 1 (2020): 7–28.
<https://doi.org/10.22503/inftars.XX.2020.1.1>
- Barnucz, Nóra, and Erna Uricska. "Innovatív nyelv tanulási módszerek és módszertan a rendészeti szaknyelvi képzés vizsgálatában. Előtanulmány egy vizsgálathoz." *Új Pedagógiai Szemle* 70, no. 9–10 (2020): 53–63.
- Barnucz, Nóra. "Book review. Uricska Erna's COVIDictionary: Words and Phrases Related to the Global Pandemic." *Papers in Art and Humanities* 1, no. 2 (2021): 43–47.
<https://doi.org/10.52885/pah.v1i2.79>

- Barnucz, Nóra, and Erna Uricska. "Kiterjesztett valóság és közösségi oldalak alkalmazása a nyelvoktatásban különös tekintettel a rendészeti szaknyelvre." *Rendvédelem* 10, no. 2 (2021): 4–48.
- Barnucz, Nóra. "Járványszótár: A világjárványhoz köthető szavak és kifejezések. COVIDICTIONARY Words and phrases related to the global pandemic." *Belügyi Szemle* 70, no. 3 (2022): 643–648.
<https://doi.org/10.38146/BSZ.2022.3.11>
- Bocsi, Veronika. "Munkaértékek a felsőoktatásban." In *Felsőoktatás és munkaerőpiac – Eseményektől a kompetenciák felé*, edited by Tibor Darvai, 67–85. Szeged: Setup Belvedere Meridionale, 2013.
- Borszéki, Judit. "E-learning anyagok használata az angol rendészeti szakmai nyelv oktatásában." *Rendőrségi Tanulmányok* 2, no. 3 (2019): 115–149.
- Boswijk, Albert, Thomas J. P. Thijssen, and Ed Peelen. *Experience Economy – a New perspective*. London: Pearson Education, 2006.
- Brown, H. Douglas. *Teaching by principles*. Englewood Cliffs, NJ: Prentice Hall, 1994.
- Chamot, Anna Uhl, and Pamela Beard El-Dinary. "Children's Learning Strategies in Language Immersion Classrooms." *The Modern Language Journal* 83, no. 3 (1999): 319–338.
- Christián, László. *Komplementer rendészet*. Budapest, Magyarország: Ludovika Egyetemi Kiadó, 2022.
- Christián, László, and Ákos Erdős. "Vészharang és jubileum? A rendészeti felsőoktatás kilátásai, a tisztjelöltek toborzásának és életpályára állításának nehézségei." *Belügyi Szemle* 68, no. 12 (2020): 11–42.
<https://doi.org/10.38146/BSZ.2020.12.1>
- Christián, László, Zoltán Hautzinger, and Gábor Kovács. "A magyar rendészeti felsőoktatás jelene és jövője." *Magyar Rendészet* 21, Special Issue (2021): 53–72.
<https://doi.org/10.32577/mr.2021.1.ksz.4>
- Czékman, Balázs. "Tablettel támogatott oktatási intézményi implementáció." In *A tanulás új útjai*, edited by Julianna Mrázik, 75–90. Budapest: Magyar Nevelés- és Oktatáskutatók Egyesülete (HERA), 2017a.
- Czékman, Balázs. "Mobiltechnológia a tanórán: oktatási tartalmak, oktatást segítő digitális megoldások." In *I. Mobil eszközök az oktatásban konferencia. Válogatott tanulmányok az I. Mobil eszközök az oktatásban konferenciáról*, edited by Péter Fehér, and Dóra Orsolya Aknai, 249–254. Debrecen: Debreceni egyetemi Kiadó, 2017b.
- Csíkszentmihályi, Mihály. *Tehetséges gyerekek. Flow az iskolában*. Budapest: Nyitott Könyvműhely, 2010.
- Csíkszentmihályi, Mihály. *Flow – az áramlat – a tökéletes élmény pszichológiája*. Budapest: Akadémiai Kiadó, 2015.
- Dani, Erzsébet. "A kétfázisú HY-DE-modell: a hiper- és mélyfigyelem fázisváltásai a katedrától a hallgatói önfejlesztésig." *Informatika a felsőoktatásban 2014 konferencia*. Accessed: May 15, 2020.
<https://bit.ly/2XWlvHI>
- Dicheva, Darina, Christo Dichev, Gennady Agre, and Galia Angelova. "Gamification in Education: A Systematic Mapping Study." *Educational Technology & Society* 18, no. 3 (2015): 1–14.

-
- Dominek, Dalma Lilla. "Élményalapú tárlatvezetés." In *Múzeumandragógia és közösségszolgálat*, edited by Mihály Kurta, Gábor Veres, and Attila Verók, 22–36. Eger: Eszterházy Károly Egyetem; Pulszky Társaság, Magyar Múzeumi Egyesület múzeumandragógiai tagozata, 2017.
- Dominek, Dalma Lilla. *Flow, avagy játékos kommunikáció a múzeumokban: Az élménycentrikus és az információcentrikus tárlatvezetés megítélése és lehetősége két hazai múzeumban*. Szeged: Belvedere Meridionale Kiadó, 2020.
- Dominek, Dalma Lilla. "Andragógia-pszichológia - FLOW típusú módszer a felsőoktatásban." In *Tanulás – Tudás – Innováció a felsőoktatásban – reflektorfényben az innováció kihívásai*, edited by Krisztina Fodorné Tóth, 209–217. Pécs: MELLearn Egyesület, 2021a.
- Dominek, Dalma Lilla. "Creativity in higher education through the flow channel." *Belvedere Meridionale* 33, no. 4 (2021b): 5–12.
<https://doi.org/10.14232/belv.2021.4.1>
- Dominek, Dalma Lilla. *Élményalapú környezeti nevelés a természettudományi múzeumokban*. Szeged: Belvedere Meridionale Kiadó, 2021c.
- Dominek, Dalma Lilla. "A flow, mint a pozitív pszichológia jelenléte az oktatásban." *Eruditio – Educatio* 21, no. 4 (2021d): 72–82.
<https://doi.org/10.36007/eruedu.2021.4.72-82>
- Dominek, Dalma Lilla, and Szabolcs Ceglédi. "Alkoss-a találékonyág pszichológiája." *Magyar Pedagógia* 121, no. 4 (2021): 395–407.
<https://doi.org/10.17670/MPed.2021.4.395>
- Dominek, Dalma Lilla. "On a Flow-based pedagogical model. The emergence of experience and creativity in education." *Eruditio – Educatio* 17, no. 3 (2022): 72–81.
<https://doi.org/10.36007/eruedu.2022.3.072-081>
- Hanushek, Erik A. "Testing, Accountability, and the American Economy." *The ANNALS of the American Academy of Political and Social Science* 683, no. 1 (2019): 110–128.
<https://doi.org/10.1177/0002716219841299>
- Institutional Development Plan, "A Biztos Jövő Egyeteme 2020-2025 - Stratégia a Harmadik Évezred Hajnalán." Accessed April 2, 2020.
<https://bit.ly/3MFAFHU>
- Kelemen, Krisztián, and János Talabér. "„Szavak nélkül a nyelv halott” – „A szóbeli kommunikáció prioritása az idegennyelv-oktatásban." In *Sokszínű pedagógiai kultúra*, edited by Judit Torgyik, 420–426. online book, 2014.
- Mathes, Jennifer. *Global quality in online, open, flexible and technology enhanced education. An analysis of strengths, weaknesses, opportunities and threats*. Norway, Oslo: International Council for Open and Distance Education (ICDE), 2019.
- Mayer, Richard E. *Multimedia learning*. New York, Cambridge: University Press, 2001.
- McQuiggan, Scott, Jamie McQuiggan, Jennifer Sabourin, and Lucy Kosturko. *Mobile Learning: A Handbook for Developers, Educators, and Learners*. Wiley, New Jersey, USA: SAS Institute Inc. 2015.
- Metropolitan Police Service, "Digital Communications Officer." Accessed January 2, 2020.
<https://policecareers.tal.net/vx/mobile-0/appcentre-External/brand-3/candidate/so/pm/6/pl/1/opp/1761-Digital-%20Communications-Officer/en-GB>
- Molnár, Katalin, and Erna Uricska. "A Café módszer alkalmazásának tapasztalatai a rendészeti felsőoktatásban." *Magyar Rendészet* 19, no. 4 (2019): 69–80.
<https://doi.org/10.32577/mr.2019.4.5>

- Nahalka, István. "Konstruktív pedagógia – egy új paradigma a láthatáron (III)." *Iskolakultúra* 7, no. 4 (1997): 22–40.
- Niyetbaeva, Gulmira, Laura Shalabayeva, Bakyt Zhigitbekova, Gulzira Abdullayeva, and Gulzhanar Bekmuratova. "Psychological and Pedagogical Conditions for Effective Application of Dialogic Communication among Teenagers." *International Journal of Environmental & Science Education* 11, no. 8 (2016): 11239–11247.
- Paivio, Allan. "Mental imagery in associative learning and memory." *Psychological Review* 76, no. 3 (1969): 241–263.
- Paivio, Allan and Alain Desrochers. "A dual coding approach to building memory." *Canadian Journal of Psychology* 34, no. 4 (1980): 388–899.
- Rhodes, Mel. "An analysis of creativity." *Phi Delta Kappan* 42, no. 7 (1961): 305–311.
- Sharma, Lata Hermant, and Poonam Gupta. "Constructivist Approach for Teaching English: Making Sense of Paradigm Shift from the Traditional Approach." *International Journal of Science and Research (IJSR)* 5, no. 10 (2016): 788–792.
- Soffer, Tal, Tali Kahan, and Rafi Nachmias. "Patterns of students' utilization of flexibility in online academic courses and their relation to course achievement." *International Review of Research in Open and Distributed Learning* 20, no. 3 (2019): 204–220.
<https://doi.org/10.19173/irrodl.v20i4.3949>
- Sung, Yao-Ting, Chang Kuo-En, and Liu Tzu-Chien. "The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis." *Computers & Education* 94, no. 2 (2016): 252–275.
- Szabó, József, and András Buda. "A távolléti oktatás tapasztalatai a Debreceni Egyetemen az oktatók körében végzett kutatások alapján." In *Tanuló társadalom: Oktatókutatás járvány idején*, edited by Rita Kattein-Pornói, Julianna Mrázik, and Mónika Pogátsnik, 259–274. Magyarország, Debrecen: Debreceni Egyetemi Kiadó, Magyar Nevelés- és Oktatókutatók Egyesülete (HERA), 2022.
- Szabó, Orsolya. "A flow élmény kapcsolatai az önmegvalósítással, a hétköznapi élet eseményekkel, a szubjektív boldogsággal és az élettel való elégedettséggel." 2014. Accessed April 02, 2022.
<https://bit.ly/3s0qlRA>
- Teng, Mark Feng, and Danyang Zhang. "The association between working memory and the effects of multimedia input on L" vocabulary learning." *International Review of Applied Linguistics in Language Teaching*. 2021.
<https://doi.org/10.1515/iral-2021-0130>
- Turner, Noleen. *What is m-learning*. 2012. Slideshare. Accessed March 25, 2016.
<http://bit.ly/294LMg0>
- Uricska, Erna. "Közösségi rendészet – közösségi oldalak: Elméleti háttér és a rendészeti digilektus fogalmának bevezetése." *Magyar Rendészet* 20, no. 2 (2020a): 153–168.
<https://doi.org/10.32577/mr.2020.2.8>
- Uricska, Erna. "Proper interactive communication of the police as a(n e-)trust-building strategy. Introducing the term policing digilect." *Kosice Security Review* 10, no. 2 (2020b): 185–195.
- Uricska, Erna. "Rendészeti közösségi oldalak alkalmazása a szaknyelvoktatásban." *Educatio* 29, no. 4 (2020c): 653–662.
<https://doi.org/10.1556/2063.29.2020.4.10>

-
- Uricska, Erna. *COVIDictionary. Words and phrases related to the global pandemic*. Magyarország, Budapest: Rejtjel Kiadó, 2021.
- Uricska, Erna. “COVIDictionary – Virolingvisztika és az angol egynyelvű járványszótár.” In *LLL 4.0 – Hogyan alakítja át a digitalizáció az LLL stratégiákat?*, edited by Krisztina Fodorné Tóth, 472–479. Pécs: MELLearn Egyesület, 2022.
- Uricska, Erna, and Katalin Molnár. “Az angol nyelvű járványszótár létrehozása szerzői és szerkesztői szemmel.” In *Tanuló társadalom. Oktatáskutatás járvány idején*, edited by Rita Kattein-Pornói, Julianna Mrázik, and Mónika Pogátsnik, 556–563. Magyarország, Debrecen: Debreceni Egyetemi Kiadó, Magyar Nevelés- és Oktatáskutatók Egyesülete (HERA Évkönyvek IX.), 2022.
- Uricska, Erna, and Máté Suták. “Közösségi oldalak rendészeti profiljain található bejegyzések alkalmazása szaknyelvoktatás során.” *Magyar Rendészet* 22, no. 1 (2022): 107–119.
<https://doi.org/10.32577/mr.2022.1.7>
- Ürmösné Simon, Gabriella, and Éva Kovács. “Az online oktatás hatékonysága a rendészettudományi karon az oktatók és a hallgatók szemüvegén keresztül – Egy hibrid kutatás eredményei.” *Védelem, Tudomány: Katasztrófavédelmi online tudományos folyóirat* 7, no. 1 (2022): 191–219.
- Wurdinger, Scott D., and Julia A Carlson. *Teaching for experiential for approaches that work*. Lanham, MD: Rowman and Littlefield Publishers, 2009.
- Wurdinger, Scott D., and Pete Allison. “Faculty perceptions and use of experiential learning in higher education.” *Journal of e-Learning and Knowledge Society* 13, no. 1 (2017): 15–26.
- Zhongyuan, C. Williams. “The Use of Multimedia Material in Teaching Chinese as a Second Language and Pedagogical Implications.” Thesis, University of Massachusetts Amherst, 2013. Accessed August 04, 2015.
<http://bit.ly/294tjA7>

Non-ICT Students' Familiarity with Basic Computer Concepts at the Faculty of Philosophy at the University of Novi Sad

As a continuation of our previous paper, the aim of this research was to determine the level of familiarity of the freshmen at the Faculty of Philosophy, University of Novi Sad, Novi Sad, Serbia, with the basic concepts related to computers. The sample included the same 568 respondents as in our first study, and the same two control groups. Familiarity was analysed with 14 questions from the field of computer literacy, both at the level of individual questions and in relation to the total number of correct answers. The results were also analysed in relation to the condition of passing the test from the appropriate European Computer Driving Licence module. No student answered all the questions correctly, which clearly indicates the need to educate students on basic concepts related to computers. The obtained results can be used to better adapt teaching to the students' needs and their prior knowledge.

Keywords: *computer and information literacy, digital literacy, students*

Acknowledgements

M. Ivanović acknowledges financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2022-14/200125).

Author Information

Zoltán Gellér, University of Novi Sad, Faculty of Philosophy, Novi Sad, Serbia

<https://orcid.org/0000-0003-1853-4740>

Dragana Milin, University of Novi Sad, Faculty of Philosophy, Novi Sad, Serbia

<https://orcid.org/0000-0003-3420-1591>

Tanja Jevremov, University of Novi Sad, Faculty of Philosophy, Novi Sad, Serbia

<https://orcid.org/0000-0003-3780-3424>

Dubravka Valić Nedeljković, Novi Sad School of Journalism, Novi Sad, Serbia

<https://orcid.org/0000-0002-9729-9752>

Mirjana Ivanović, University of Novi Sad, Faculty of Sciences, Novi Sad, Serbia

<https://orcid.org/0000-0003-1946-0384>

How to cite this article:

Gellér, Zoltán, Dragana Milin, Tanja Jevremov, Dubravka Valić Nedeljković, Mirjana Ivanović. "Non-ICT Students' Familiarity with Basic Computer Concepts at the Faculty of Philosophy at the University of Novi Sad". *Információs Társadalom* XXIII, no. 2 (2023): 25–47.

<https://dx.doi.org/10.22503/inftars.XXIII.2023.2.2>

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

Today, both in literature and in educational practice, the need for digital or computer literacy as the most important “functional” literacy is no longer questioned and such literacy is now considered essential in modern society. Today, “digital competences, computer skills, information literacy and related abilities represent a crucial element in ICT education (Information and Communication Technologies)” (Stopar and Bartol 2019, 479). Many authors, including Bernd W. Becker (2018), still debate whether digital and information literacy are the same, or whether digital literacy (DL) has emerged as a separate consideration out of information literacy. After considering different approaches to this debate, Becker concluded that “digital literacy and information literacy cannot exist without one another” and that “we might consider grouping students and patrons by how ready they are to embrace digital literacy skills” (Becker 2018, 2).

Many authors define DL in a similar way. Spante et al. stated that “In more recent publications, definitions of DL point towards cognitive skills and competences” (Spante et al. 2018, 7). This indicates how much the topic of defining the field covered by the term “digital literacy” is still essential and relevant. In the present paper, we single out two prototypical definitions: one where “...digital literacy emphasizes reading, writing, understanding, evaluating, communicating, and using the information in different formats” (Dewi, Pahriah, and Purmadi 2021, 88), and another where “Digital Literacy is the awareness, attitude and ability of individuals to appropriately use digital tools and facilities to identify, access, manage, integrate, evaluate, analyse and synthesize digital resources, construct new knowledge, create media expressions, and communicate with others, in the context of specific life situations, in order to enable constructive social action; and to reflect upon this process” (Martin and Grudziecki 2006, 255). These definitions indicate that DL is a very complex set of skills and knowledge, and that approaches to mastering those skills must be based on clear insights into the level of freshmen’s prior knowledge, which should be identified by empirical research. Indeed, only digitally literate students, as many authors have noted, can be successful students in study programmes in various disciplines of higher education today, including social-humanistic fields.

A specific level of DL as such is not an end in itself. Cadiz-Gabejan and Takenaka showed in their study how important DL is for students’ overall success in studying: “The findings of the study revealed that the students needed to enhance the extent of their computer literacy in the areas of word processing, spreadsheet, presentation, and general computing. The results also signified that the greater the extent of their computer literacy in said areas, the higher their academic performance” (Cadiz-Gabejan and Takenaka 2021, 29). They indicated how important computer training courses are for students by stating that “an intervention program is imperative to enhance the extent of the students’ computer literacy, especially that they felt some constraints with it” (Cadiz-Gabejan and Takenaka 2021, 42).

Another heavily debated subject is the tension “between the theoretical conceptualizations of DL as a multidimensional construct and empirical studies reporting unidimensional DL scores. Also, little is known about how DL may vary among dif-

ferent age cohorts, and whether and at which age do performance gaps emerge with respect to gender” (Jin et al. 2020, 1).

In recent years, there has been a lot of empirical research on the level of DL among different target groups. In a previous paper (Gellér et al. 2021), we reported on the familiarity of freshmen at the Faculty of Philosophy at the University of Novi Sad, Novi Sad, Serbia, with the basic concepts of the Internet. This follow-up paper focuses on their prior knowledge in basic computer literacy. The presented results are based on the same analysis carried out at the beginning of the winter semester of 2019 as in our previous study.

Research carried out by Rizal et al., which used a very similar methodology and target group to ours, showed that “the mean of student’s digital literacy was 50 with a low category. Three digital literacy competency areas owned by students show that the mean literacy competency of information and digital data is 36 (low category), the mean of communication and collaboration competency is 68 (medium category), and the mean of digital content creation is 47 (low category)” (Rizal et al. 2020, 1).

Reddy et al. focused on an identical target group as in our research in their study. In an effort to assess the DL of first-year students at a regional university in the South Pacific, they concluded that “...the students who join higher education are digitally literate. However, students are not competent in all aspects of digital literacy. Therefore, educators need to design and develop appropriate interventions and training programs which comprise all aspects of digital literacy” (Reddy et al. 2020, 5).

Another research study on the DL of students led by Hina Amin confirmed that “digital literacy is a multi-dimensional construct and requires a comprehensive theoretical background that should include all major elements of digital literacy that one should possess to thrive in the digital world” (Amin, Abid Malik, and Akkaya 2022, 37). This research also tested students’ DL competency and knowledge based on a “Digital Literacy Scale (DLS) based on Chen’s (2015) theoretical framework which includes nine dimensions: communication, collaboration, critical thinking, creativity, citizenship, character, curation, copyright, and connectedness” (Amin, Abid Malik, and Akkaya 2022, 24). Their conclusion was that this technique is very useful for testing the level of DL not only among students but also the general population.

The study presented by Tham et al. considered the way students think about their digital knowledge and skills, stating that “student perceptions of their own digital literacy (and instructor’s awareness of such) are informed by prior experiences by means of metaphors and mental models. These conceptualizations can shape how they learn with technology” (Tham et al. 2021, 14). In concluding their paper, the authors recommended that “Instructors should incorporate a variety of digital technologies into coursework that allow students to practice multidimensions of digital literacy” (Tham et al. 2021, 15).

All the above-mentioned reports spoke about the need to empirically investigate the level of students’ knowledge in the field of DL, which was also the subject of our study, to verify the opinion that the new generations of students are already sufficiently digitally literate by themselves and that they do not need additional courses in that field to succeed in their studies.

The rest of the paper is organised as follows: the next section describes the study sample and the methodology used, while the third section is dedicated to a detailed

analysis and statistical comparison of the students' answers, and finally the last section presents a discussion of the results.

2. Study Sample and Methodology

The study aimed to analyse the familiarity of freshmen at the Faculty of Philosophy (FF) at the University of Novi Sad, Novi Sad, Serbia, with basic computer concepts. The study sample comprised 522 students from 13 out of 17 study programmes (subgroups) at the FF (programmes with a small number of students were not included). The research also included two control groups: one from the Faculty of Technical Sciences (FTN), comprising 29 students, and one from the Faculty of Management (FAM), comprising 17 students. The total number of student participants in the study sample was thus 568. The distribution of students by study programmes is shown in Figure 1. For a clearer presentation of the results, instead of the full names of the (sub) groups, we henceforth use their abbreviations given in Figure 1.

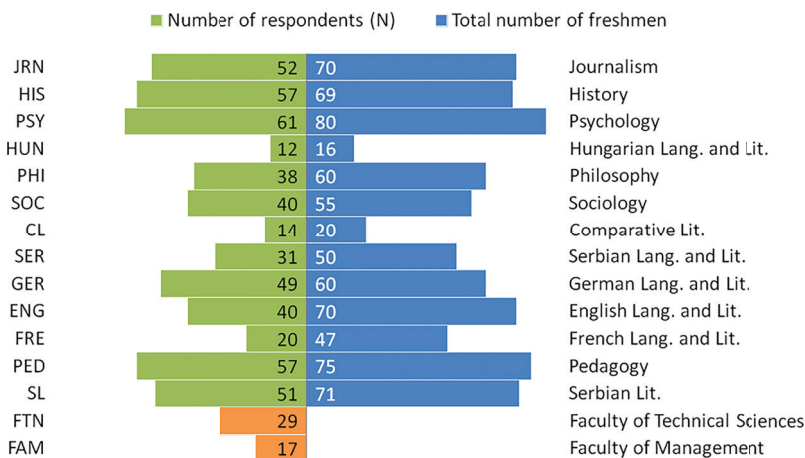


Figure 1. Study programmes and the numbers of respondents. (self-edited)

The knowledge test applied here was based on a questionnaire composed of 22 close-ended questions, with 4 answers possible per question. The questions were selected from a set of questions used in the final tests in the subject of *Computer Literacy* at the FF. Here, 8 questions referred to the basic concepts of the Internet and 14 questions to computer-related concepts. We presented an analysis of students' answers to the questions related to the basic concepts of the Internet in our earlier study (Gellér et al. 2021). The current paper focuses on the analysis of the students' answers to the 14 questions related to the use of computers.

The testing was conducted anonymously; the students had 15–20 minutes to complete the test in a printed paper form.

2.1. Data analysis

The familiarity of the freshmen of the FF with basic computer concepts was analysed both at the level of the individual questions and in relation to the total number of correct answers, and was compared with the results from the control groups. The scores of the FF students as a whole (i.e. for all the subgroups together) were examined, as well as the scores of the individual subgroups too.

The percentages of different answers to the individual test questions are shown graphically. Pearson's χ^2 test and Cramer's V coefficient were used to assess the statistical significance of the connections among the different faculties and individual subgroups within the FF based on their numbers of correct answers. Distributions of correct answers, measures of central tendency, and the variability among the control groups and individual FF subgroups were examined as well. The students' results were analysed in relation to the requirements of passing the test from the appropriate European Computer Driving Licence (ECDL) module (Computer Essentials) and in relation to how many of them would get a negative grade on the exam at the respective faculties.

Examination of differences in the results of the students of the FF and the control groups was done using the Kruskal–Wallis test, with multiple comparisons of the average ranks performed using the Bonferroni-adjusted z-test.

The interconnectedness of the individual test questions with respect to the number of correct answers was checked by calculating the phi correlation coefficients between the questions.

3. Results

The presentation of the research results starts with a review of the questions and answers, with graphical displays showing the shares of the different answers to the individual test questions. The abbreviations in parentheses after the ordinal numbers of the questions are later used in the tables and charts to increase their readability. Note, "X" in the figures refers to students who "did not answer the question".

3.1. Overview of the students' answers

Q1 (OS). One of the basic tasks of operating systems is to manage computer resources, such as the RAM, processor, input, and output devices. The first question tested whether first-year students recognise this task. When asked: "What is the software called that manages computer resources?", the students were offered the following answers (note, "3. operating system" is the correct answer):

1. application software
2. antivirus program
3. operating system
4. BIOS

Figure 2 shows that the vast majority (80.8%) of first-year students at the FF answered this question correctly and that they were more successful than both the control groups: FTN (72.4%), FAM (41.2%).

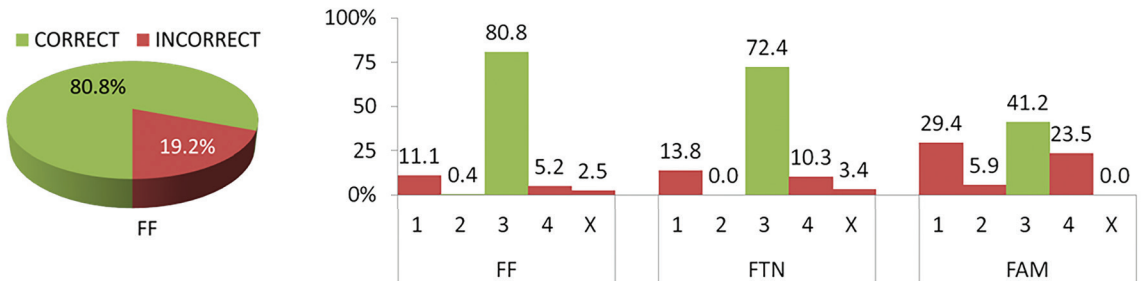


Figure 2. Q1 (OS) – share of different answers (self-edited)

Among the wrong answers, the most common one was “application software” in all three groups: FF (11.1%), FTN (13.8%), and FAM (29.4%).

Q2 (EXE). The general form of file names is NAME.EXT, where the extension determines the type of file and indicates its purpose. In the Windows OS, programs are stored in executable files with the EXE extension. The aim of the second question was to determine whether the respondents could recognise the correct extension for executable files among the following options (note, “2. EXE” is the correct answer):

1. DAT
2. EXE
3. HTML
4. PNG

Less than half of the FF students chose the correct answer (47.9%, Figure 3). The general-purpose extension DAT, which does not belong to any well-known file category, was chosen by almost one-quarter of the students (24.9%), and the HTML extension representing web pages by one-fifth (20.9%). The PNG extension, which relates to images, was chosen by only 18 out of the 522 FF students (i.e. 3.4%).

In terms of the control groups, there were visible differences in the student responses. While a large percentage of the FTN respondents (82.8%) answered the question correctly, the percentage of correct answers among the FAM students (17.6%) was significantly lower than among the FF students. More than half of the FAM group students (58.8%) put DAT as the correct answer.

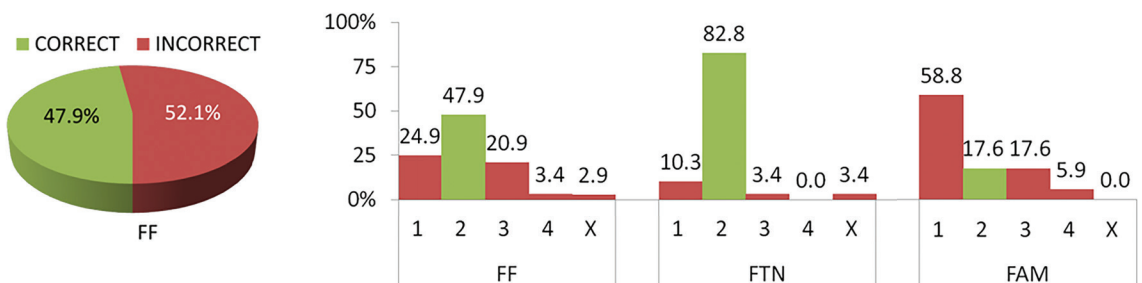


Figure 3. Q2 (EXE) – share of different answers (self-edited)

Q3 (RAM). The third question examined whether the freshmen could recognise the acronym that stands for the main memory in a computer (note, “4.RAM” is the correct answer). The following options were offered:

1. GPU
2. ROM
3. CPU
4. RAM

GPU stands for Graphics Processing Unit and refers to the graphics processor in charge of creating the images that are displayed on a computer monitor. ROM stands for Read Only Memory and refers to the non-volatile memory in a computer whose contents cannot be changed, only read. CPU stands for the Central Processing Unit and refers to the central processor of a computer that executes programs loaded into its main memory (RAM: Random Access Memory) whose content can be changed, not just read as in the case of ROM.

Based on Figure 4, it is clear that recognising the acronym of the main memory did not present a problem for the vast majority of students (FF: 86.4%, FTN: 86.2%, FAM: 100%). Moreover, the FF students achieved the best result for this question among all the questions.

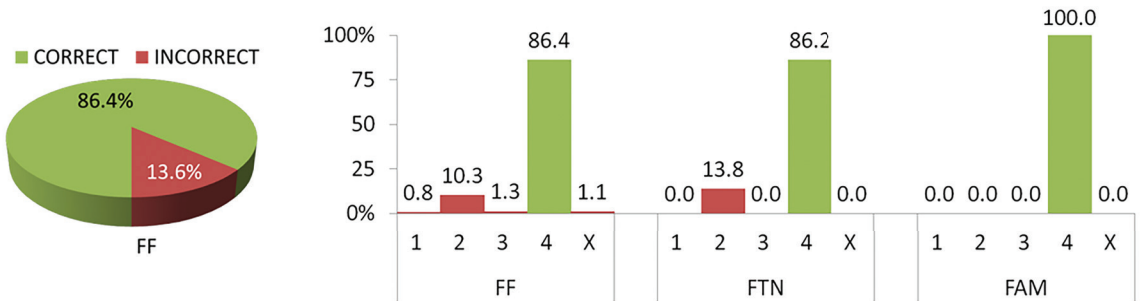


Figure 4. Q3 (RAM) – share of different answers (self-edited)

Q4 (KGT). As knowledge of units of measurement is crucial in so many fields, including when working with files and folders, the fourth question was focused on the names and relationships between the standard units in computing. The task for the students here was to choose the correct sorting (in ascending order) of the names of the data measures (note, “1. kilobyte, gigabyte, terabyte” is the correct answer):

1. kilobyte, gigabyte, terabyte
2. megabyte, kilobyte, gigabyte
3. byte, bit, kilobyte
4. gigabyte, terabyte, kilobyte

When it came to the FF students, based on Figure 5, it can be seen that the correct sorting of data measures was a problem for 40.8% of the respondents. Almost one-quarter of the students (23.8%) mistakenly thought that a kilobyte was a larger unit than a megabyte, and 12.1% of them did not notice the correct relationship between bit and byte.

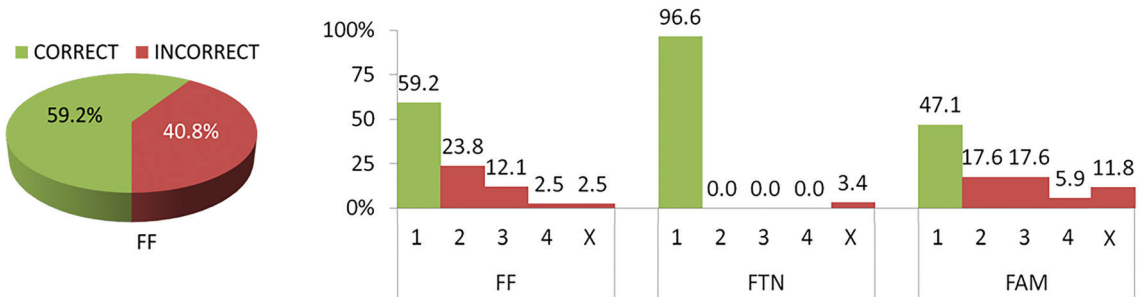


Figure 5. Q4 (KGT) – share of different answers. (self-edited)

For comparison, in the FTN group, all the students who answered this question chose the correct answer, while less than half of the respondents in the FAM group selected the correct sorting (47.1%).

Q5 (PC). In modern society, most people use various smart devices, especially smartphones and tablets. With the fifth question we wanted to examine whether students know that tablets belong to the family of personal computers. When asked what a tablet is, the following answers were offered (note, “1. personal computer” is the correct answer):

1. personal computer
2. super computer
3. mainframe computer
4. a program that loads before the operating system

From Figure 6, it can be seen that more than half of the respondents from all three faculties chose the correct answer. The percentage of correct answers was the highest in the FAM group (76.5%), but the lowest in the FTN group (51.7%), with the FF group in between these (65.7%).

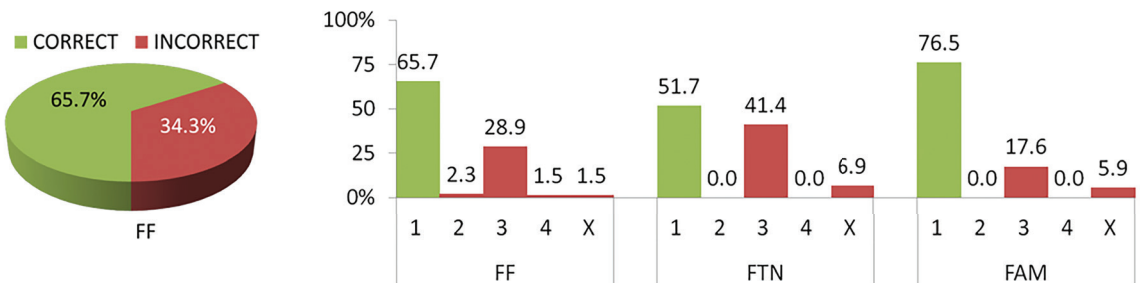


Figure 6. Q5 (PC) – share of different answers (self-edited)

It was also evident that the students who answered this question incorrectly almost exclusively identified tablets as being associated with mainframe computers. However, they could clearly distinguish the term supercomputer from the term tablet.

Q6 (BS). When naming files and folders, special care must be taken to avoid using characters that have a special role within the file system (like “/”, which is separator in the OS) and their use is generally prohibited. For instance, the file management program that is an integral part of the Windows OS does not allow the usage of “/” in file and folder names. However, not all programs are so strict and some of them allow the use of “/” when naming files, which can lead to unexpected results. With this in mind, we asked students why the name “Domaći zadatak 2014/2 Excel.7z“ (Homework 2014/2 – Excel.7z) is not allowed for Windows files. The possible answers were as follows (note, “4. contains a slash “/”” is the correct answer):

1. the extension must be XLS or XLSX
2. contains spaces
3. contains the letter “ć”
4. contains a slash “/”

In the example name in the question, 7z is a legal extension that indicates a compressed file, while the spaces and the letter “ć” are allowed to be used in both file names and folder names.

The results in Figure 7 show that more than two-thirds (71.6%) of the FF students either did not recognise the use of “/” as the reason why the stated name was not allowed for files or did not answer the question, with 36.2% of the respondents thinking that the problem was in the file extension, 17.6% an error in the spaces, and 14.4% the use of the letter “ć”.

In the control groups, the percentage of correct answers was even lower: 17.2% (FTN) and 23.5% (FAM). A significantly higher percentage of FTN students (55.2%) noted the extension as the reason for the bad file name compared to respondents from the FF (36.2%) and the FAM (35.3%) groups.

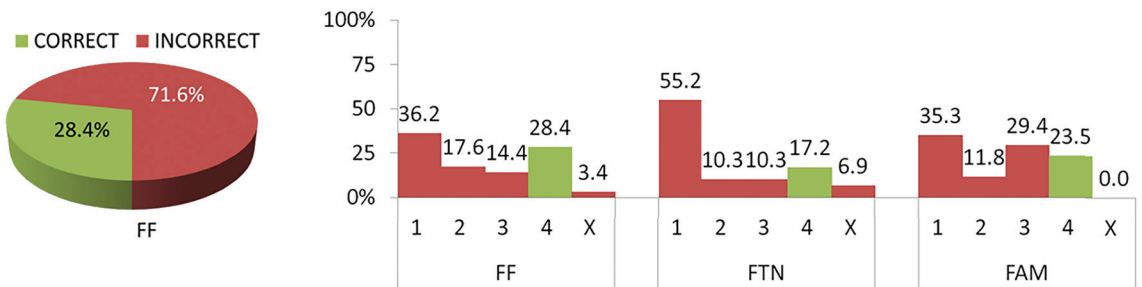


Figure 7. Q6 (BS) – share of different answers (self-edited)

Q7 (MB). This question investigated whether the freshmen knew which of the following parts of the computer connects its various components and enables communication between them (note, “3. motherboard” is the correct answer):

1. processor
2. network card
3. motherboard
4. RAM memory

The task of the processor is to execute programs loaded into the RAM of the computer, while the network card allows communication between networked computers.

The percentage of students in the FF group who answered correctly was 50.2% (Figure 8), which was slightly less than in the FAM group (52.9%) and significantly less than in the FTN group (69%).

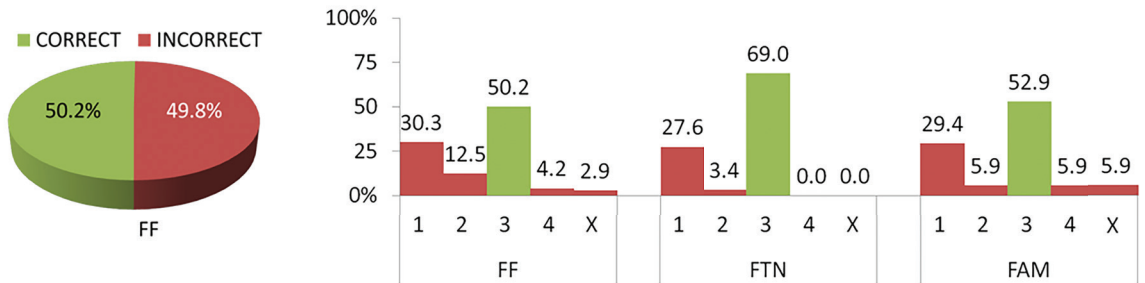


Figure 8. Q7 (MB) – share of different answers (self-edited)

Nearly one-third (30.3%) of the respondents in the FF group marked the processor as the correct answer. This was also the most common among the incorrect answers in the control groups.

Q8 (SCR). In this question, we were interested in whether students knew what the screen resolution is (note, “1. number of points on the width and height of the monitor” is the correct answer):

1. number of points on the width and height of the monitor
2. number of open windows
3. number of different colours on the screen
4. the size of the icons on the screen

Based on Figure 9, it can be seen that the percentage of correct answers was very high, both among the FF students (85.2%) and in the control groups: 93.1% (FTN) and 76.5% (FAM).

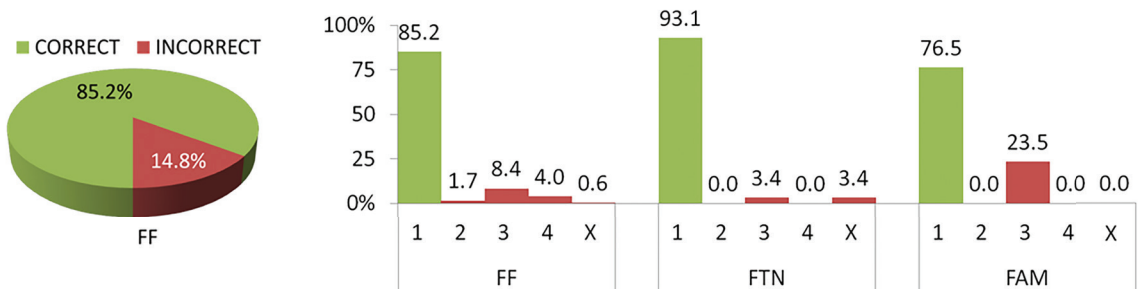


Figure 9. Q8 (SCR) – share of different answers (self-edited)

Q9 (IN). The ninth question asked respondents to identify input devices among the following pairs of computer components (note, “1. scanner and microphone” is the correct answer):

1. scanner and microphone
2. printer and speakers
3. sound and graphics card
4. processor and motherboard

Figure 10 shows that many of the FF students confused the terms input and output devices: most of them (34.5%) chose the second option. There was a significant number of those who considered the third option gave the input devices (76 out of 522, i.e. 14.6%) and almost every tenth (50 out of 522, i.e. 9.6%) respondent chose option 4. Also, the percentage of students who did not answer this question was the highest among all the questions (7.9%, i.e. 41 out of 522).

While the majority of the students in the FTN group accurately identified the input devices (75.9%), close to two-thirds of the FAM group chose output devices instead of input devices (64.7%).

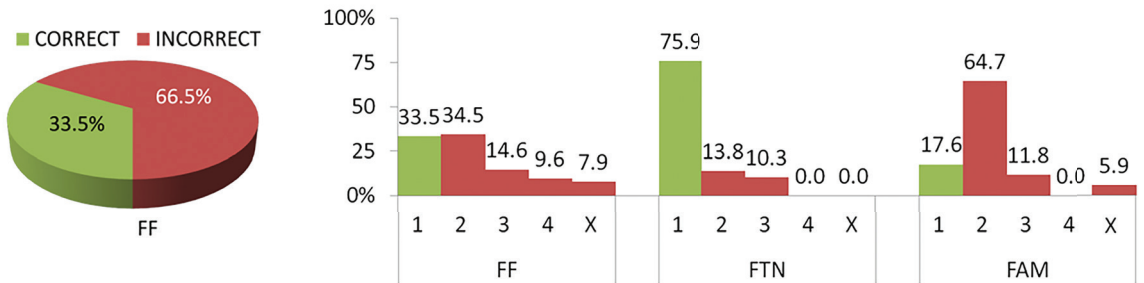


Figure 10. Q9 (IN) – share of different answers (self-edited)

Q10 (DIR). At the top of the hierarchical structure of the file system are storage devices, on which data are stored in files organised into folders. Folders can contain files and other folders called subfolders. The tenth question was about what folders can contain (note, “4. files and folders” is the correct answer):

1. only files
2. only subfolders
3. discs and files
4. files and folders

As can be noticed from Figure 11, this question was not a problem for most students. The percentage of correct answers was the highest in the FAM group (100%), followed by the FTN group (96.6%), and finally the target group of this study (FF, 86%).

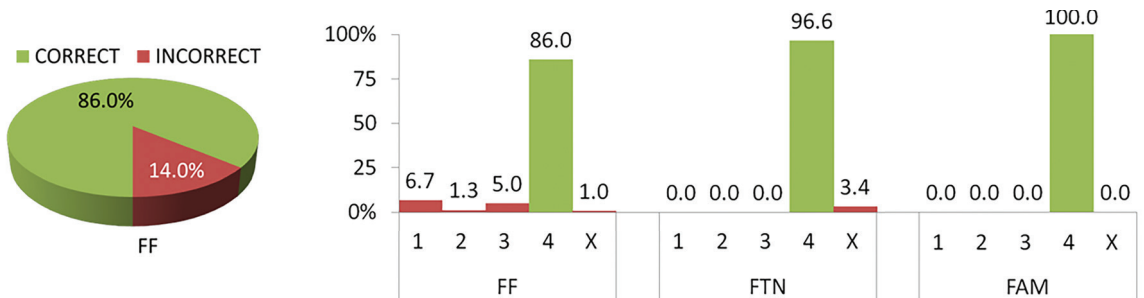


Figure 11. Q10 (DIR) – share of different answers (self-edited)

Q11 (SBY). By applying the “Sleep” command, a computer will enter standby mode with reduced power consumption, while programs and data remain in the main memory. By moving the mouse or pressing a key on the keyboard, the computer will wake up and resume normal operation. When asked what the “Sleep” command allows, the possible answers were (note, “2. activating standby mode” is the correct answer):

1. locking the computer
2. activating standby mode
3. shutting down the computer
4. starting a drawing program

Just over the half of the FF students gave the correct answer (55.7%, Figure 12), while 28.4% (148 out of 522) of respondents in the FF group chose the first option, and 12.5% (65 out of 522) thought that the “Sleep” command was used to turn off the computer.

The percentage of correct answers was much higher in both the control groups: 86.2% (FTN) and 70.6% (FAM).

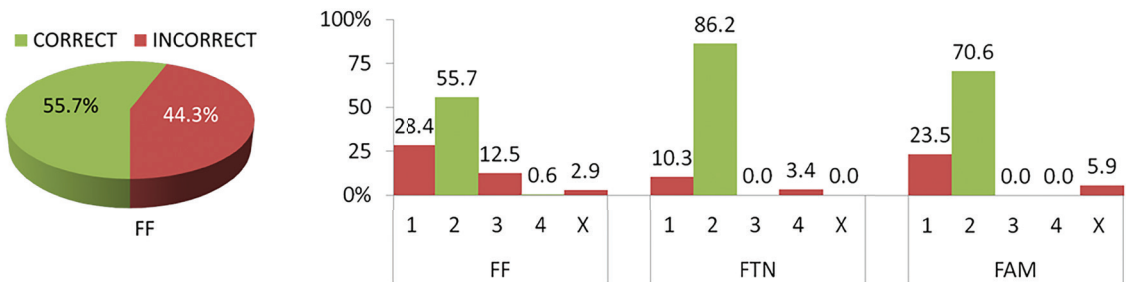


Figure 12. Q11 (SBY) – share of different answers (self-edited)

Q12 (APP). Application software includes programs that are designed for specific user needs, such as word processing, image processing, spreadsheets, audio and video playback, and others. The task of the respondents in this question was to choose which of the following programs is application software (note, “3. Photoshop” is the correct answer):

1. Linux
2. BIOS
3. Photoshop
4. Android

Here, Photoshop is a well-known image processing program, while Linux and Android are operating systems, and BIOS (Basic Input / Output System) is the computer basic input output system.

Figure 13 indicates that recognising the application software presented a difficulty for the FF students, and more than two-thirds (70.7%) of respondents answered incorrectly. Indeed, more students (37.2%) chose Android than those who answered correctly (29.3%). One in five respondents (20.5%) chose the Linux OS, while many others (9.4%, i.e. 49 out of 522) opted for BIOS.

In the FTN group, the percentage of correct answers was visibly higher (48.3%), but the percentage of those who chose Android was also slightly higher (41.4%) compared to the results of the FF students. In this group, only two respondents (out of 29) selected Linux as the answer, while no-one chose BIOS.

The percentage of correct answers was the lowest in the FAM group (23.5%), in which the share of students who did not answer the question was the highest (23.5%, i.e. 4 out of 17 students). It is also interesting to note that Android was chosen by only one respondent in this group and that the percentages of responses related to Linux and BIOS were higher than in both the FTN and FF groups.

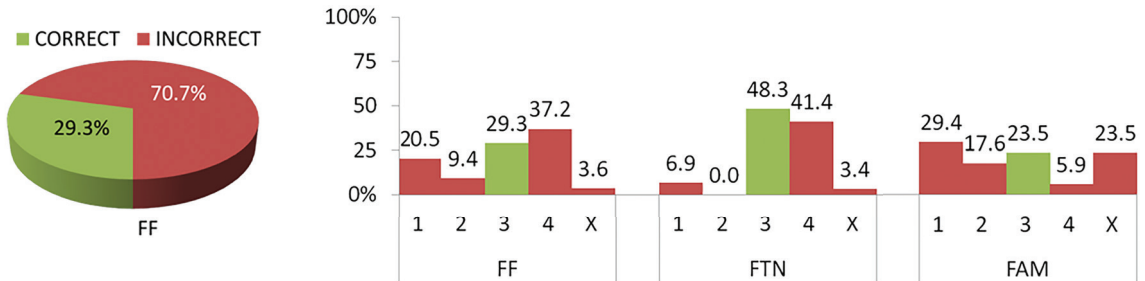


Figure 13. Q12 (APP) – share of different answers (self-edited)

Q13 (SHR). Software refers to programs that run on computers, while hardware refers to the physical components of a computer. Programs that can be used for free without restrictions are known as freeware, while shareware represents programs that can be used for free but have a time and/or functional limitation. In this question, we asked students to choose the category of programs that can be used for free with some restrictions (note, “4. shareware” is the correct answer):

1. software
2. freeware
3. hardware
4. shareware

This question seems to be one of the most difficult for all the students to answer, regardless of faculty. None of the respondents in the control groups answered this question correctly (Figure 14), while in the target group (FF), only approximately every tenth answer was correct (11.1%).

The majority of respondents chose the category of programs that can be used free of charge without any restrictions (Freeware).

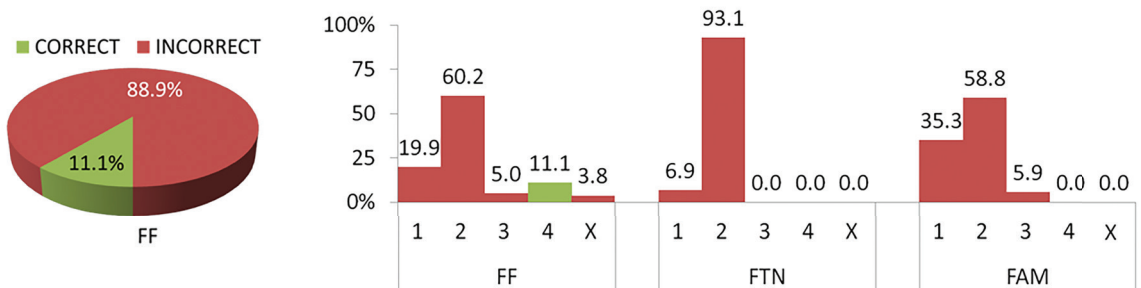


Figure 14. Q13 (SHR) – share of different answers. (self-edited)

Q14 (FN). With the last question, we wanted to examine whether the students could recognise the Hungarian scientist John von Neumann, who is credited with setting up the basic architecture of modern computers. The alternative answers were other well-known names in computer science and informatics, specifically Steve Jobs (one of the founders of Apple), Mark Zuckerberg (co-founder of the Facebook), and Alan Turing (one of the fathers of computer science) (note, “4. John von Neumann” is the correct answer):

1. Steve Jobs
2. Mark Zuckerberg
3. Alan Turing
4. John von Neumann

It can be seen from Figure 15 that the name of John von Neumann was not known among most the FF students. As many as 40.8% of them indicated that the architecture of modern computers originated from Steve Jobs, while approximately one-third thought that it was Alan Turing.

Similar results were observed in the FAM group, where Alan Turing was picked by slightly fewer students (23.5%), while the percentage of students who thought that basic computer architecture came from the founder of Facebook corresponded to the same percentage of students who marked John von Neumann as the correct answer (17.6%).

In the FTN group, 20 out of 29 respondents answered correctly (69%), while Steve Jobs was chosen by 4 students (13.8%), and Mark Zuckerberg and Alan Turing by 2 each (6.9%).

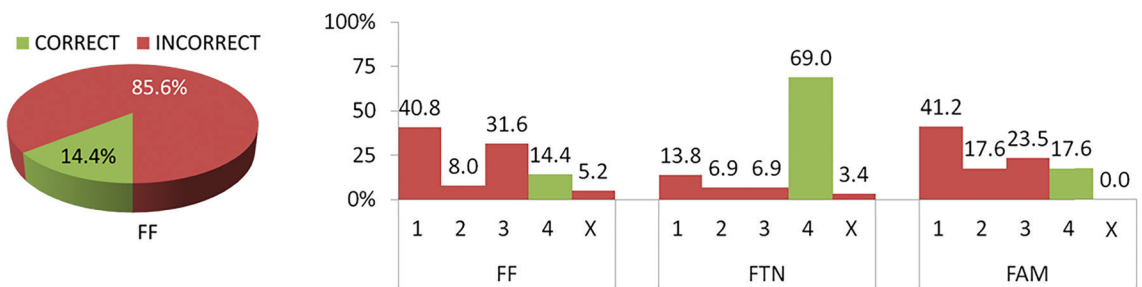


Figure 15. Q14 (FN) – share of different answers. (self-edited)

3.2. Percentages of correct answers

An overview of the percentages of correct answers by study programmes (rows) and by questions (columns) is given in Table 1.

The best results of the study programmes are marked in bold, and the weakest are underlined. For example, students of journalism (JNR) achieved the best result in the case of the SCR question (92.3%), and the weakest in the case of the SHR and FN questions (9.6%). The cumulative results of the FF students are shown in the row

FF, and the cumulative results of all the participating students are given in the row marked as ALL.

	OS	EXE	RAM	KGT	PC	BS	MB	SCR	IN	DIR	SBY	APP	SHR	FN	TEST
JRN	75.0	50.0	80.8	69.2	59.6	17.3	40.4	92.3	26.9	88.5	48.1	30.8	<u>9.6</u>	<u>9.6</u>	49.9
HIS	63.2	50.9	84.2	59.6	57.9	22.8	66.7	82.5	29.8	84.2	54.4	17.5	<u>5.3</u>	19.3	49.9
PSY	85.2	59.0	96.7	59.0	78.7	32.8	52.5	83.6	50.8	93.4	68.9	36.1	16.4	<u>14.8</u>	59.1
HUN	91.7	33.3	83.3	66.7	91.7 ●	41.7	25.0 ○	66.7 ○	25.0	83.3	83.3	8.3 ○	<u>0.0</u> ○	25.0	51.8
PHI	71.1	50.0	89.5	52.6	68.4	21.1	44.7	92.1	34.2	89.5	55.3	42.1	<u>10.5</u>	21.1	53.0
SOC	87.5	45.0	92.5	47.5	65.0	27.5	47.5	77.5	42.5	80.0	50.0	27.5	<u>12.5</u>	<u>12.5</u>	51.1
CL	78.6	28.6	92.9	57.1	57.1	14.3 ○	50.0	78.6	28.6	100 ●	64.3	42.9	<u>0.0</u> ○	7.1	50.0
SER	83.9	38.7	90.3	71.0	51.6 ○	45.2 ●	51.6	93.5	22.6	90.3	58.1	19.4	<u>9.7</u>	12.9	52.8
GER	77.6	42.9	85.7	61.2	55.1	32.7	44.9	85.7	36.7	85.7	59.2	20.4	18.4	<u>6.1</u> ○	50.9
ENG	87.5	85.0 ●	80.0	77.5	70.0	40.0	70.0 ●	95.0 ●	65.0	92.5	70.0	50.0 ●	<u>12.5</u>	<u>12.5</u>	64.8
FRE	95.0 ●	55.0	65.0 ○	70.0	55.0	<u>15.0</u>	40.0	90.0	25.0	90.0	65.0	20.0	20.0 ●	25.0	52.1
PED	91.2	36.8	82.5	38.6 ○	71.9	28.1	40.4	73.7	<u>14.0</u> ○	71.9 ○	43.9	21.1	15.8	15.8	46.1 ○
SL	80.4	29.4	90.2	56.9	72.5	29.4	54.9	88.2	23.5	82.4	39.2 ○	37.3	<u>2.0</u>	13.7	50.0
FTN	72.4	82.8	86.2	96.6 ●	51.7	17.2	69.0	93.1	75.9 ●	96.6 ●	86.2 ●	48.3	<u>0.0</u> ○	69.0 ●	67.5 ●
FAM	41.2 ○	17.6 ○	100 ●	47.1	76.5	23.5	52.9	76.5	17.6	100 ●	70.6	23.5	<u>0.0</u> ○	17.6	47.5
FF	80.8	47.9	86.4	59.2	65.7	28.4	50.2	85.2	33.5	86.0	55.7	29.3	<u>11.1</u>	14.4	52.4
ALL	79.2	48.8	86.8	60.7	65.3	27.6	51.2	85.4	35.2	87.0	57.7	30.1	<u>10.2</u>	17.3	53.0
MIN	41.2	17.6	65.0	38.6	51.6	14.3	25.0	66.7	14.0	71.9	39.2	8.3	0.0	6.1	46.1
MAX	95.0	85.0	100	96.6	91.7	45.2	70.0	95.0	75.9	100	86.2	50.0	20.0	69.0	67.5

Table 1. Percentages of correct answers (self-edited)

Among the terms covered by the test, the FF students were best acquainted with the main memory (RAM, 86.4%), folders (DIR, 86%), screen resolution (SCR, 85.2%), and operating system (OS, 80.8%), while they had the most difficulties with recognising shareware (SHR, 11.1%), the name of the scientist who laid the foundation of modern computer architecture (FN, 14.4%), the use of the slash as an illegal character when naming files and folders (BS, 28.4%), application software (APP, 29.3%), and computer input devices (IN, 33.5%).

The best results by question in Table 1 are marked by the symbol ●, while the symbol ○ indicates the worst. For example, the best result in the case of the first question (OS) was for the FRE subgroup (95%), and the weakest for the FAM subgroup (41.2%). Summary results for all the test questions are given in the last column of the table, marked as TEST. In this category, the FTN group (67.5%) was the most successful, and the least successful was the PED subgroup (46.1%).

From the graphical representations of the percentages of correct answers in Figure 16, it is easier to see the mutual relations between the questions, individual study programmes, and different faculties.

The results for the FTN group were visibly better than for the other (sub)groups in the case of the EXE, KGT, IN, SBY, and FN questions (Chart A). Students in the HUN subgroup achieved significantly better results than the other subgroups in the case of the PC and SBY questions, but were noticeably worse in the case of the MB and APP questions. The FAM group was noticeably worse than the other (sub)groups in answering the OS and EXE questions, yet all the students in that group answered the

RAM and DIR questions correctly. The percentages of correct answers by the ENG subgroup were among the highest in the case of several questions: EXE, KGT, MB, SCR, IN, and APP.

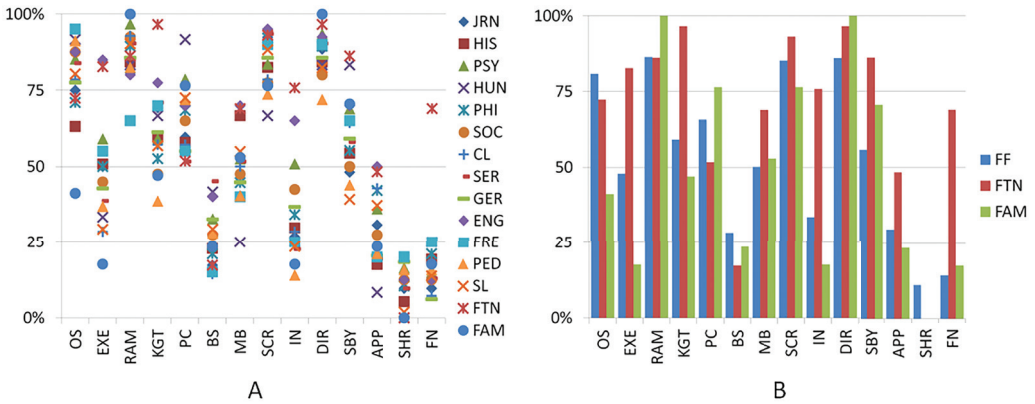


Figure 16. Graphical representation of the percentage of correct answers by study programme (A) and faculty (B) (self-edited)

The biggest differences among the accuracies of the group responses were observed in the recognition of the executable file extension (EXE) and computer input device (IN) terms. Looking at the performances per question, by ignoring 1–2 atypical values, two groups of questions with homogeneous results could be observed. The first group included questions with more than 70% correct answers by groups: the main memory (RAM), screen resolution (SCR), and folders (DIR). The second group consisted of questions with the lowest percentage of correct answers (less than 40%): slash as an example of an illegal character when naming files and folders (BS), shareware (SHR), and John von Neumann (FN).

Figure 16B indicates that the ratio of the percentage of correct answers of the students from different faculties varied from question to question.

The FTN group stood out for its significantly better results in identifying the executable file extension (EXE), proper sorting of data measures (KGT), application software (APP), and the role of John von Neumann in modern computing (FN). However, when asked what a tablet (PC) is and why the file name is incorrect (BS), they clearly gave less correct answers than the FF and FAM students.

The students in the FAM group achieved better results than the students in both the FTN and FF groups in recognising the acronyms of the main memory (RAM), determining what a tablet (PC) is, and choosing what folders can contain (DIR). However, their responses were clearly worse than those of the FTN and FF groups when it came to knowing the role of the OS and the file name extension EXE.

The FF students were slightly more successful than the students in both control groups in determining the OS roles, the cause of the file name incorrectness (BS), and the category of programs that can be used for free with some restrictions (SHR).

3.3. Statistical significance of the differences in the answers among the (sub)groups

Table 2 shows the results from testing the differences in the number of correct answers among the students of different study programmes within the FF (A) and among the students of different faculties, i.e. FF, FTN, and FAM (B).

Considering the level of significance $\alpha=0.05$, the students in the FF study programmes differed in their success of answering most questions (OS, EXE, KGT, MB, SCR, IN, SBY, and APP). Significant differences were found between the students in the FF, FTN, and FAM in terms of their answers to the OS, EXE, KGT, IN, SBY, and FN questions.

As expressed by Cramer's V coefficient, the relationships between the test performances and study programmes were relatively mild in strength, even with the results that proved significant for interpretation.

The accuracy in answering the RAM, PC, BS, DIR, and SHR questions was not significantly related neither to the study programmes at the FF nor to the study programmes of the other two faculties. As can be seen in Table 1 and Figure 16, below are the questions that most students, regardless of study programme, answered correctly (RAM, PC, SCR, and DIR), or incorrectly (BS and SHR).

	A			B		
	df=12, N=522			df=2, N=568		
	χ^2	p	V	χ^2	p	V
OS	26.107	0.010	0.224	16.601	0.000	0.171
EXE	40.426	0.000	0.278	20.160	0.000	0.188
RAM	20.324	0.061	0.197	2.667	0.264	0.069
KGT	23.924	0.021	0.214	17.453	0.000	0.175
PC	19.703	0.073	0.194	3.334	0.189	0.077
BS	17.213	0.142	0.182	1.844	0.398	0.057
MB	22.273	0.035	0.207	3.896	0.143	0.083
SCR	21.181	0.048	0.201	2.475	0.290	0.066
IN	43.882	0.000	0.290	23.956	0.000	0.205
DIR	19.254	0.083	0.192	5.317	0.070	0.097
SBY	23.389	0.025	0.212	11.631	0.003	0.143
APP	27.947	0.006	0.231	5.056	0.080	0.094
SHR	17.098	0.146	0.181	5.692	0.058	0.100
FN	10.102	0.607	0.139	57.366	0.000	0.318

Table 2. Results from the χ^2 tests and Cramer's V coefficients analyses of the relationships among the answer correctness and the study programmes at the FF (A) and among the included faculties (B) (self-edited)

3.4. Distributions of answers

Figure 17 shows the indicators of the central tendency and variability of the cumulative results on the tests for the individual study programmes at the FF, and for the FTN and FAM control groups.

Students of different FF programmes mostly performed slightly worse than the FTN students but better than the FAM students, getting approximately half the answers correct. ENG proved to be the most successful FF subgroup, in which, as in the FTN group, 50% of students had 9 or fewer correct answers, and the other 50% more than that. This was followed by the PSY subgroup, in which half the students successfully answered 8 or fewer questions, while the other half achieved more than that.

In the HUN, SER, GER, and FRE subgroups, the arithmetic mean was shifted to slightly lower values, as a probable consequence of the existence of less frequent lower results. Less successful were the PED subgroup, with about 6.5 correct answers on average, and the SL subgroup, with 50% correct answers to 6 or fewer questions.

None of the students, including those from the most successful (sub)groups, answered all the questions correctly.

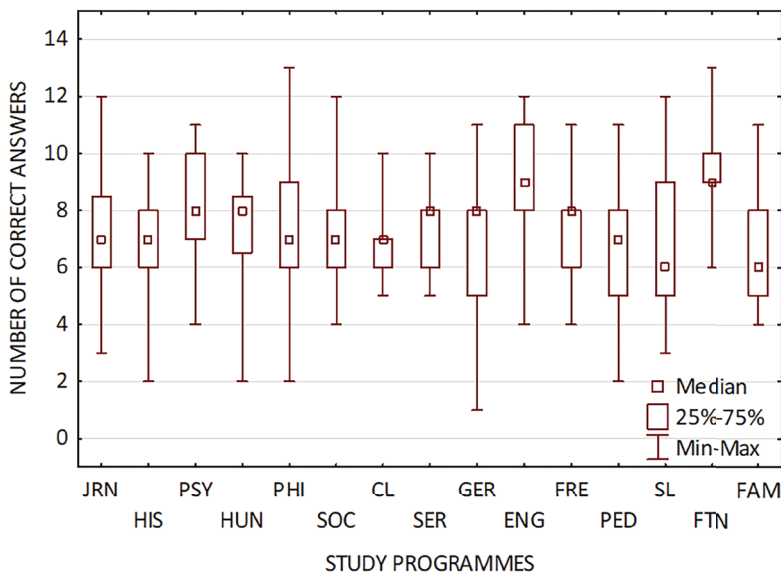


Figure 17. Measures of the central tendency and variability of the correct answers (self-edited)

Apart from the differences in average and mean values, the distributions of the correct answers of the individual (sub)groups also differed in terms of their variability.

In the relatively small CL subgroup, in which the largest number of students got about half of the answers correct, the differences in performance were small. In the

other small HUN subgroup, however, individual differences in the number of correct answers were significantly larger.

The variability was low in the CL and FTN (sub)groups. The ENG and PSY subgroups, which proved to be approximately similarly successful as the FTN group, had increased variability due to there being a smaller number of low scores, and the distributions were somewhat asymmetric. A similar asymmetry characterised the HIS and HUN subgroups.

The Kruskal–Wallis test showed that the described differences in results among the students of the individual FF, FTN, and FAM groups could be considered statistically significant ($H(14, N=568)=83,894, p<0.001$). The applied z-tests showed that this result was a consequence of a higher number of correct answers by the FTN students compared to most of the students of the other (sub)groups; the more correct answers of the ENG students in relation to the SOC, GER, SL, HIS, JRN, FAM, and PED students; and the higher number of correct PSY student responses compared to those of the PED students.

3.5. Student’s results in relation to the ECDL standard and the requirements for passing the exam

To get a clearer picture of how familiar the students were with the basic concepts of computing, we compared the number of correct answers by study programme, focusing on how well they would meet the test requirement from the appropriate ECDL module (Computer Essentials) and how many of them would fail the faculty exam (Figure 18).

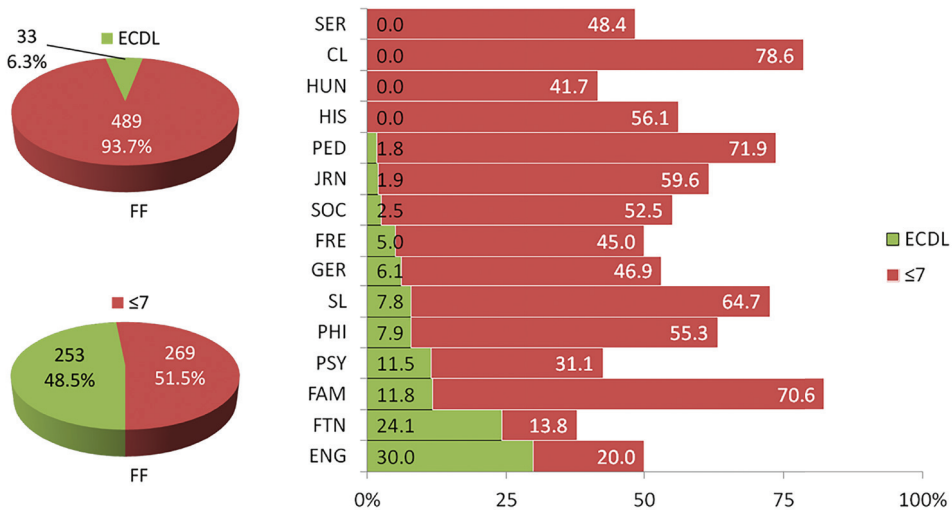


Figure 18. Percentage of students who meet the condition defined by the ECDL standard and answered correctly less than 51% of the questions (≤7) (self-edited)

Successful passing of the ECDL test requires 75% correct answers from each candidate. As our test contained 14 questions, this meant it was necessary to answer at least 11 questions correctly to pass the test. Out of a total of 522 FF freshmen who undertook the test, only 33 met this condition (6.3%). When the individual study programmes were observed, no-one in the SER, CL, HUN, and HIS subgroups reached this limit. In the other subgroups, the percentage of respondents who answered more than 10 questions correctly was less than 10%, with the exception of the PSY (11.5%, i.e. 7 out of 61) and ENG (30%, i.e. 12 out of 40) subgroups. According to this criterion, the most successful subgroup was ENG.

A positive grade requires at least 51% correct answers on the exam, which in the case of our test corresponded to 8 correct answers. Figure 18 shows that about half of the FF freshmen would have failed the exam (269 out of 522 students). For the individual study programmes, the percentages ranged between 20% (8 out of 40) in the ENG subgroup and 78.6% (11 out of 14) in the CL subgroup. While the best result was achieved in the FTN group (13.8%, i.e. 4 out of 29), the percentage of FAM group students who would fail the exam was among the highest, amounting to 70.6% (12 out of 17).

3.6. Correlations between the questions

Correlations between the accuracy of the answers to the questions on the total sample, expressed by phi coefficients showed very low correlations between the accuracies of the answers to the questions (Table 3). There was no strong regularity that if a student answered one of the questions correctly that she or he would then answer the other questions correctly as well. Somewhat higher though were the correlations between the following pairs of questions: EXE and KGT, EXE and SBY, RAM and DIR, and KGT and SBY. The heights of these correlations, however, did not exceed 0.22.

	OS	EXE	RAM	KGT	PC	BS	MB	SCR	IN	DIR	SBY	APP	SHR	FN
OS		0.039	0.044	-0.030	-0.054	-0.023	-0.005	0.046	0.050	0.021	0.001	0.100	0.044	0.004
EXE			-0.056	0.193	0.001	0.090	0.156	0.045	0.144	0.085	0.200	0.097	0.043	0.030
RAM				0.049	0.033	0.055	0.140	0.030	0.026	0.204	0.077	0.007	0.028	-0.001
KGT					-0.056	0.054	0.081	0.116	0.178	0.117	0.217	0.127	-0.003	0.062
PC						-0.062	-0.015	-0.019	-0.044	0.026	-0.084	-0.022	-0.060	-0.078
BS							-0.003	0.022	0.080	0.052	0.074	-0.011	-0.013	0.041
MB								0.105	0.151	0.104	0.164	0.180	0.027	-0.002
SCR									0.107	0.092	0.050	0.119	0.074	0.004
IN										0.099	0.138	0.127	0.007	0.024
DIR											0.092	0.037	-0.042	0.094
SBY												0.072	0.088	0.089
APP													0.070	0.005
SHR														-0.031
FN														

Note: Bold values are significant at the level 0.05

Table 3. Correlation coefficients among the individual questions for the total sample of respondents (self-edited)

A similar pattern of relationships among the answers was present within the largest observed group, i.e. the FF students.

Within the observed subsamples of FF and FAM students, the number of correlations that were significant for interpretation was smaller, due to the lower probability that the indicator would be significant with fewer measurements.

In the FTN group, a positive correlation was detected between the following pairs of questions: SCR and DIR, SBY and EXE; whereby, none of the students answered the SHR question correctly, but due to the lack of data variability, it was not possible to calculate the correlations between this question and the others.

There were no correct answers to the SHR question in the FAM group either; however, all of the students in that group answered the RAM and DIR questions correctly. In this group, a negative correlation was observed between the OS and SBY questions.

The higher correlations and their significance obtained in the FTN and FAM groups were possible consequences of random variations that are more likely with small samples.

4. Discussion

The assumption that today's generation growing up in the information age adopt knowledge in the field of computer and information literacy "on the go" and that they do not need formal education was not proved to be justified here. The total percentage of correct answers from FF students was 52.4%, which was only slightly better than the total percentage of their correct answers to questions in the field of Internet knowledge, which was 49.6% (Gellér et al. 2021).

Bearing in mind that the ECDL standard requires 75% of the questions to be correctly answered to obtain a certificate, only 33 (6.3%) of the FF freshmen would meet this criterion. More than one-third of these (12) were from the ENG subgroup, and there were some study subgroups (HIS, HUN, CL, and SER) in which no candidate achieved the required level of success. This result was significantly worse than the results related to the basic concepts of the Internet, where almost 20% of the FF participants would have met the ECDL requirement (Gellér et al. 2021).

The question that turned out to be the easiest for all students to answer was the one related to the objects that can be contained in folders (with a success rate of 87%), while the majority of FF students also recognised the acronym that denotes main computer memory (RAM) (86.4%). The most difficult question for all the students was to choose the program category that can be used free of charge with some restrictions (SHR) (11.1% among the FF students, and 10.2% of all students overall).

The students from the FTN control group performed better when it came to those questions that were "more technical" in nature (executable file extension (EXE), correct sorting of data measures (KGT), application software (APP), and the role of John von Neumann in modern computing (FN)), which could be explained by the prior knowledge of the students and their affinities that influenced their choice of faculty. What was surprising was the much worse answers to questions that were "more

user” focused in nature, such as what is a tablet (PC) and why is the given file name incorrect (BS), because this group gave less accurate answers than the other (sub) groups of students.

The FF students showed somewhat worse results than those from the FTN group and somewhat better results than the ones from the FAM group. The most successful FF subgroup was ENG followed by PSY. No student of the FTN had less than 6 correct answers, and 50% of their scores between the first (Q1) and third (Q3) quarters were placed in the range between 9 and 10 points. In contrast, in the ENG and PSY subgroups, the minimum number of correct answers was 4 and the interquartile range was higher, which indicated that the knowledge of the FTN students was more homogeneous. Also, it is worth nothing that the two best-performing FF subgroups were the ones who had the most registered candidates for the entrance exam (and thus the strictest selection), and their results were likely a consequence of their better success during their high school education.

The fact that none of the 568 students answered all the questions correctly also underlines the need to better educate students on basic computer concepts.

References

- Becker, Bernd W. “Information Literacy in the Digital Age: Myths and Principles of Digital Literacy.” *School of Information Student Research Journal* 7, no. 2 (2018).
<https://doi.org/10.31979/2575-2499.070202>
- Cadiz-Gabejan, Alona Medalia, and Melinda Jr C. Takenaka. “Students’ Computer Literacy and Academic Performance.” *Journal of World Englishes and Educational Practices* 3, no. 6 (2021): 29–42.
<https://doi.org/10.32996/jweep.2021.3.6.4>
- Dewi, Citra Ayu, Pahriah Pahriah, and Ary Purmadi. “The Urgency of Digital Literacy for Generation Z Students in Chemistry Learning.” *International Journal of Emerging Technologies in Learning (IJET)* 16, no. 11 (2021): 88.
<https://doi.org/10.3991/ijet.v16i11.19871>
- Gellér, Zoltán, Tanja Jevremov, Karlo Bala, Dubravka Valić Nedeljković, and Mirjana Ivanović. “Non-ICT Students’ Familiarity with Basic Internet Services and Tools at the Faculty of Philosophy at the University in Novi Sad.” *Információs Társadalom* 21, no. 4 (2021): 38.
<https://doi.org/10.22503/inftars.XXI.2021.4.3>
- Amin, Hina, Muhammad Abid Malik, and Bulent Akkaya. “Development and Validation of Digital Literacy Scale (DLS) and Its Implication for Higher Education.” *International Journal of Distance Education and E-Learning* 7, no. 1 (2022): 24–43.
<https://doi.org/10.36261/ijdeel.v7i1.2224>
- Jin, Kuan-Yu, Frank Reichert, Louie P. Cagasan, Jimmy de la Torre, and Nancy Law. “Measuring Digital Literacy across Three Age Cohorts: Exploring Test Dimensionality and Performance Differences.” *Computers & Education* 157 (November 2020): 103968.
<https://doi.org/10.1016/j.compedu.2020.103968>

- Martin, Allan, and Jan Grudziecki. "DigEuLit: Concepts and Tools for Digital Literacy Development." *Innovation in Teaching and Learning in Information and Computer Sciences* 5, no. 4 (2006): 249–67.
<https://doi.org/10.11120/ital.2006.05040249>
- Reddy, Pritika, Kaylash Chaudhary, Bibhya Sharma, and Ronil Chand. "Digital Literacy: A Catalyst for the 21 St Century Education." In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1–6. Gold Coast, Australia: IEEE, 2020.
<https://doi.org/10.1109/CSDE50874.2020.9411548>
- Rizal, Rahmat, Dadi Rusdiana, Wawan Setiawan, and Parsaoran Siahaan. "The Digital Literacy of The First Semester Students in Physics Education." *Jurnal Pendidikan Fisika* 8, no. 2 (2020): 101–10.
<https://doi.org/10.26618/jpf.v8i2.3293>
- Spante, Maria, Sylvana Sofkova Hashemi, Mona Lundin, and Anne Algers. "Digital Competence and Digital Literacy in Higher Education Research: Systematic Review of Concept Use." *Cogent Education* 5, no. 1 (2018): 1519143.
<https://doi.org/10.1080/2331186X.2018.1519143>
- Stopar, Karmen, and Tomaž Bartol. "Digital Competences, Computer Skills and Information Literacy in Secondary Education: Mapping and Visualization of Trends and Concepts." *Scientometrics* 118, no. 2 (2019): 479–98.
<https://doi.org/10.1007/s11192-018-2990-5>
- Tham, Jason Chew Kit, Kenyan Degles Burnham, Daniel L. Hocutt, Nupoor Ranade, John Misak, Ann Hill Duin, Isabel Pedersen, and Jessica Lynn Campbell. "Metaphors, Mental Models, and Multiplicity: Understanding Student Perception of Digital Literacy." *Computers and Composition* 59 (March 2021): 102628.
<https://doi.org/10.1016/j.compcom.2021.102628>

Can team communication obscure severe usability problems?

Communication analysis of teams using shared visual workspaces

Team communication is a significant topic of study in the psychological literature, but it is a less researched area in relation to software usability. In this article, as part of a larger research project, we investigated team communication to explore the relationship between usability problems and team communication patterns. The research question was examined in two laboratory experiments as part of a collaborative software evaluation process. For sequential analysis of team communication transcripts, we created a custom code system based on previous literature. The results confirm that teams that experience special types of usability problem show different communicational patterns from teams that experience no such problems. The results also show high reliability of the new code system. Further research is needed to explore the relationship between usability problems and team communication patterns in different settings.

Keywords: *Human-computer interaction, collaborative software, communication analysis, lag-sequential analysis, team usability testing*

Author Information

Dalma Geszten, Budapest University of Technology and Economics

<https://orcid.org/0000-0003-4336-0674>

Balázs Péter Hámornik, Budapest University of Technology and Economics

<https://orcid.org/0000-0001-8085-1259>

How to cite this article:

Gesztén, Dalma, and Balázs Péter Hámornik. "Can team communication obscure severe usability problems?". *Információs Társadalom* XXIII, no. 2 (2023): 48–66.

==== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.3> ====

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

How users use different technology in their everyday lives is a key topic in the human-computer interaction research field (Tófalvy 2019; Fehér 2017; Vajda 2020; Szűts and Yoo 2018; Hámornik et al. 2014; Hercegi et al. 2015). Among several types of software, this paper focuses on collaborative software, the goal of which is to support teamwork. The first collaborative software emerged in the 1950s and became widespread from the 1990s as a consequence of the appearance of the World Wide Web and due to the popularity of teamwork as a research topic (Schmidt and Bannon 2013). Therefore, collaborative software has quickly become an integral part of everyday work, making usability a key issue.

The study presented in this article is related to the investigation of team communication in a collaborative software usability context, as part of a larger research project. Although the goal of the research project is to develop a new method for the usability evaluation of collaborative software, here the focus is on the communication analysis of teams, as it is an element of the new method, called Team Usability Testing (Geszten, Hámornik and Hercegi 2021).

First, we present the theoretical background of the topic and the broader research context. Then we describe the steps of the code system design and the results of analysing the communication of teams while using shared online visual workspaces.

2. Related literature

2.1. Teamwork: definition of team and virtual team

Teamwork is a major research topic in work psychology, as people work more effectively in a team than individually. There are many definitions related to team, the different elements of which are: being a system of interdependent individuals in which members have different roles (Hackman 1987); being characterised by the fact that its members see themselves as a social unit (Cohen and Bailey 1997); and having cooperation as a common goal (Salas, Burke and Cannon-Bowers 2000).

Globalisation, the explosion of the Internet, the development of technology, the resulting demand for rapid product development and the competitive market responses to this have all been key drivers for the widespread adoption of virtual teams (Aldag and Kuzuhara 2015). This was also facilitated by the rapid spread of collaborative software, which took place in the decade following the appearance of the World Wide Web in 1989 (i.e., 1989–99) (Schmidt and Bannon 2013).

In defining virtual teams, the continuum theories seek not to distinguish strictly between virtual and non-virtual teams but rather to view virtuality as a continuum, a characteristic of the team (Hertel, Geister and Konradt 2005; Mesmer-Magnus et al. 2011).

Virtuality refers to whether team members prefer to be face-to-face, hybrid (with some members face-to-face and others virtual) or fully virtual when working together.

Virtual teamwork has several benefits at both individual and organisational level, as well as a number of challenges, as Table 1 describes.

	Benefits	Challenges
At individual level	Flexibility (in time and space) Higher motivation Increased responsibility and authority of members	Loneliness Reduction in personal relationships Role conflicts Conflicts of interest
At organisational level	Can form teams based on expertise, not location (can gather experts on a single topic or different topics) Work can be continuous, because of different time zones Flexibility and responsiveness to changing market needs Reduced travel and office-related costs	Difficult to monitor performance Can have high cost for the right technology Data security Additional development costs Designing efficient workflows Reduced motivation among members Maintaining trust and team spirit

Table 1. Benefits and challenges of virtual teamwork (Hertel, Geister and Konradt 2005; Krumm et al. 2016)

2.2. Team communication analysis

Achieving performance beyond individual effort is one of the key goals of teamwork, and several methods have been developed to investigate the factors that influence the performance and the characteristics of a successful team (Aldag and Kuzuhara 2015; Levi and Askay 2020). Team communication analysis is one of these methods.

The characteristics of a team can be divided into three main groups:

1. cognitive (cohesion, how well they know each other)
2. affective (group mood)
3. behavioural characteristics (coordination, cooperation, communication).

Cognitive and affective factors are not clearly manifested in behaviour and are therefore more difficult to measure than behavioural characteristics. One of the most dominant behavioural characteristics of a team is communication. Communication analysis is relevant because when a team performs a cognitive task (not a physical task), communication reveals the team-level cognitive processing of the task (Cooke et al. 2012).

One definition of team communication is the exchange of information among team members (Adams 2007); indeed, in teamwork sharing information is the most important function of communication among team members. In addition, communication is an opportunity to increase motivation towards a common goal and to

provide control over tasks and responsibilities. It also has an emotional function, allowing members to express emotions and thereby reducing frustration (Aldag and Kuzuhara 2015).

Depending on how it is operationalised, team communication can mean many things, referring to the frequency, quality, duration or content of communication (Marlow, Lacerenza and Salas 2017). Team research most often investigates the frequency or the quality of communication. While frequency refers to the extent of communication, quality refers to the accuracy and clarity of communication among team members. In this study, team communication patterns are coded using our newly developed coding system and then examined with sequence analysis.

2.3. Technology as a key element of virtual team performance

As emphasised throughout the theoretical introduction, technology is a key determinant of virtual team performance. A prominent element in the various definitions of virtual teams is technology, especially the frequency and the mode of technology in use (Martins, Gilson and Maynard 2004; Kirkman and Mathieu 2005). Throughout the life cycle of virtual teams, the selection of appropriate technology is emphasised at the preparatory stage, as supported by the input-process-output model, which mentions ‘technology’ as an important input factor among the work environment factors (Saunders 2000). In addition, it highlights the task-technology relationship as a team process that also influences team performance.

Having adequate technological support for virtual teams is therefore crucial. However, providing a high enough level of technological support for a virtual team to work effectively can be very expensive, so it is essential to choose the right tools and software.

2.4. Collaborative software and its usability

2.4.1. Collaborative software

In the broadest sense, collaborative software is computer technology that helps a team collaborate using digital media (Khoshafian and Buckiewicz 1995; Yen et al. 1999). Another definition is that ‘groupware systems support collaborative work of users that share common objectives’ (Salomón et al. 2019, 11).

Collaborative systems (not just software) can be grouped in many ways. One of the most common approaches is based on where and when the interaction among participants takes place (Bafoutsou and Mentzas 2002). In terms of the temporal dimension, the interaction can take place at the same time (synchronous) or at different times (asynchronous), and in terms of the spatial dimension, it can take place in the same place (co-located/face-to-face) or in different places (remote) (Bafoutsou and Mentzas 2002).

The Team Usability Testing that is the subject of the broader research (detailed in Section 2.5.) aims to investigate the usability of collaborative software used at the same time, that is, synchronously.

A typical example of this type of collaborative software is any multi-user shared visual workspace, such as an online whiteboard. In this context, collaborative actions involve several users working together at the same time in a shared online space (e.g., editing or organising some content together).

2.4.2. Understanding the usability of collaborative software

Collaborative software usability is defined as the extent to which collaborative software enables teamwork to be carried out – efficiently, effectively and satisfactorily – for a given team’s particular joint activity (Pinelle, Gutwin and Greenberg 2003). Problems with collaborative software can be divided into two groups:

- contextual problems caused by organisational and social factors
- problems resulting from inadequate support of collaboration mechanics (Steves et al. 2001).

Use of collaborative software involves two different types of work:

1. taskwork – single-user actions required to complete a task, without involving other team members;
2. teamwork – group actions required to complete a task, performed by working together (Pinelle and Gutwin 2002; Pinelle, Gutwin and Greenberg 2003).

It is important to distinguish between these two different types of work because, while individual software evaluation methods focus only on task-related work, collaborative software usability evaluation methods need to assess support for both task- and team-related work.

2.5. The broader context of our study

In the broader context of our research, we focus on developing a new usability testing method that aims to investigate the usability of collaborative software used synchronously, that is, at the same time. The new Team Usability Testing differs from previous methods in that it is an empirical method for evaluating synchronous (real-time) collaborative software, involving real or potential users to explore usability problems. In addition to individual usability problems, team usability problems must also be considered for collaborative software. Single-user usability methods are not suitable for investigating team usability problems because these problems can only be observed in a collaborative situation. Team Usability Testing can also be used to test teams working together in the same (face-to-face) or different (remote) locations.

Team Usability Testing consists of questionnaires, screenshots and interviews. Data analysis is based on the theory of team processes and the mechanics of collaboration and consists of the analysis of communication transcripts, interviews and

questionnaire data (Marks, Mathieu and Zaccaro 2001; Pinelle, Gutwin and Greenberg 2003).

As team communication analysis is an element of Team Usability Testing, we performed a sequence analysis on the communication transcripts of the first and second lab studies using a self-developed code system. Although this paper focuses on the team communication analysis results, the development process and the usability results of the Team Usability Testing method are discussed in detail in other papers (Geszten, Hámornik and Hercegi 2018, 2019, 2021, 2023).

3. Methodology

3.1. Research questions of the first and second laboratory studies

We designed the first and second laboratory studies to investigate the feasibility of the Team Usability Testing under controlled laboratory conditions. The method aims to test the usability of a given collaborative software and to identify team usability issues related to the software, involving real/potential users.

The research question we are focusing on in this paper is: What is the relationship between usability problems and a team's communication patterns?

3.2. Participants

Ten teams participated in the first lab study: two teams in the pilot study and eight in the final study. We needed the two pilot teams to finalise the task instructions and the technical conditions. In the final study, we excluded one team from the data analysis, so the remainder of the description of participants refers to the seven teams that participated. During the final study, we assigned participants to one of two roles: collaborator or observer. Each team consisted of three collaborators; if there were four members in the team, one member became an observer. Therefore, not all teams had observers. The observers were there more for organisational reasons, to monitor the teamwork and to act as a kind of back-up, standing in for any absent collaborators, so that we could carry out the study even if not all collaborator participants came (as was the case many times).

The teams were randomly formed based on the dates of the research. Students were allowed to apply for different dates, and everyone was put in a team with the participants who applied for the same date. Participants in the collaborative role were undergraduates aged 21–28 (mean = 23.57) who knew each other. Most participants described themselves as more of a team player (on a scale of 1–7: 1 – I prefer to work alone; 7 – I prefer to work in a team) with a mean of 4.61 across teams, ranging from 3.67 to 5). Further, 9 out of 21 participants had previous experience with PREZI and all teams had at least one participant who had used it before.

Eleven teams took part in the second laboratory study. As in the first lab study, we assigned the participants one of two roles: collaborator or observer. Each team consisted of three collaborators and an observer. (In the second lab study, we invited four participants to each time slot, with the aim of making the study feasible if someone did not come. As in the first lab study, the participants drew their roles from an envelope. Since all four participants came on each occasion, there were observer participants in each case.)

In the second laboratory study, 10 men and 23 women, aged 18–22, participated in the collaborator role (mean: 19.42, standard deviation: 1.27). In the before-task questionnaire, we asked participants, ‘Do you prefer working alone or in a team?’ (on a scale of 1 to 7: 1 – mostly alone; 7 – mostly in a team). The combined mean for all teams was 4.00, with standard deviation 1.54. In the second lab study, none of the participants had previous experience with the Miro collaborative whiteboard software.

3.3. The evaluated software

The first lab study was related to PREZI, while the second lab study was related to the Miro shared visual online workspace. We chose these for two reasons. First, they were relatively new and popular in Hungary at the time of the research, but still had huge potential for development. Second, they did not require any special prior knowledge to use, even for a first-time user.

PREZI is an infinite canvas (in the words of its creators, a ‘zoomable, canvas-based editor’) that allows users to develop creative presentations in an online interface. In addition to individual work, PREZI also offers the possibility for teamwork, and can therefore be considered collaborative software (Laufer, Halácsy and Somlai-Fischer 2011).

Miro (miro.com) is a collaborative whiteboard software that provides a common visual workspace for collaborators, mainly to visualise different ideas and (work) processes.

Several users can edit the same content at the same time in a shared workspace in both PREZI and Miro. The collaborative features supported by them are related to collaborative editing: users are able to work in a shared workspace at the same time, while seeing what other users are doing and where they are in the interface.

3.4. The procedure of the test

Figure 1 shows the process of the first and second laboratory studies. More detail can be found in Geszten, Hámornik and Hercegfí (2021).

The three collaborator participants were tasked to prepare a joint PREZI presentation in 30 minutes. Their task was to organise a company team-building event and to create a presentation from their ideas. This type of task was chosen because it does not require any ‘special’ skills. In addition, the short time frame justified the

low complexity of the task. To solve the task, participants had to use the collaborative software PREZI.

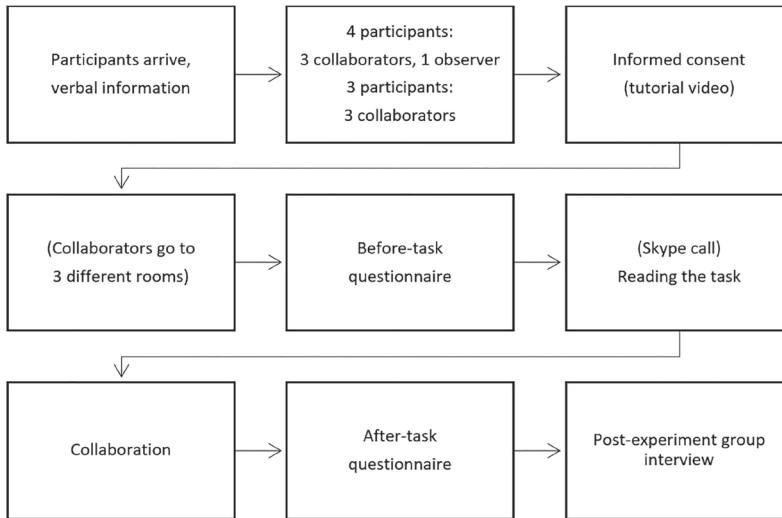


Figure 1. The procedure of the first and the second laboratory studies (in parentheses are the parts that occurred only in the first lab study due to the simulation of virtual collaboration) (created by the authors)

The procedure for the second lab study was the same as for the first lab study, the main difference being that the participants worked together face-to-face, using Miro.

3.5. Tools

The three collaborator participants worked on laptops during the task. The task was video and audio recorded using a free screen recording software OBS (Open Broadcast Software). The pre- and post- questionnaire was in Google Forms format. The group interview was recorded using a dictaphone and an audio recording application.

4. Results: analysis of the communication patterns of the teams of the first and second laboratory studies

Based on the results of the first and second lab study, we observed different usability problems in the teams that affected their collaboration. While some team usability problems occur in some teams, they do not occur in other teams (Geszten, Hámornik and Hercegfí 2021).

Among the factors affecting collaboration, we selected the most severe problem, overwriting, for further analysis. Overwriting means that one participant accidentally (due to inadequate support of collaborative functions) overwrites or deletes the work of another participant in the shared workspace. Therefore, we examined the nature of the differences in the communication patterns of the teams in which overwriting occurred, comparing to the teams in which it did not.

In the Team Usability Testing, the group interaction among team members, which in this study is the transcript of the communication, can be interpreted as a time series, that is, sequence data. ‘Group interaction is a series of messages that influence subsequent group interaction and/or reflect underlying rules of interaction such as phases that sequentially structure group interaction’ (Hewes, Poole and Hollingshead 2012, 358).

The basis of group interaction methods is that the interaction can be interpreted as a temporal pattern. This makes it possible to identify different patterns of interaction and to test hypotheses about the patterns (Hewes, Poole and Hollingshead 2012).

The two main method types are sequential contingency analysis and phasic analysis. Of the two methods of sequential contingency analysis (Markov models, lag-sequential analysis), we used lag-sequential analysis in this research. We performed the analysis using Brian O’Connor’s freely available SPSS script (O’Connor 1999).

Lag-sequential analysis examines the pattern of sequential dependencies between encoded communication acts (currently a unit of communication) in terms of conditional probability. Its main question is the probability that an event is followed by another event. If this probability is less/greater than 50%, there is a sequential relationship between events (Hámornik 2013; Hewes, Poole and Hollingshead 2012).

4.1. Design of the code system

For the sequence analysis of team communication patterns, we created a new code system, as summarised in Table 2. To develop the code system, we combined two existing coding systems: team processes (Marks, Mathieu and Zaccaro 2001) and the mechanics of collaboration (Pinelle, Gutwin and Greenberg 2003). The main codes of the code system are based on team processes theory, which contains more general, higher-level codes. Under the main codes of team processes, the collaboration mechanics have been assigned as subcodes.

Based on the code system, one of the authors has coded the communication transcripts of the teams participating in the first and second lab studies. The first three transcripts were also coded by an expert independent of the research, who was trained by the authors. Based on the reliability analysis of the coding system (later we will write about it in detail), the reliability of the coding system is high ($\kappa=0.819$; $p<0.001$). In this paper the phrases code, coding, code system and code reliability are related to social sciences. We assessed reliability by the value of the Cohen’s kappa index (Keszei et al. 2019). When examining the codes one by one,

each code is reliable. Of all the codes, only the MA (mission analysis) code has a low reliability value (below 0.5 is poor reliability).

Code name based on team processes theory (Marks, Mathieu, and Zaccaro 2001)	Code explanation	Subcode name based on the mechanics of collaboration theory (Pinelle, Gutwin and Greenberg 2003)	Code abbreviation
Mission analysis, formulation, planning	Evaluating and interpreting the team's purpose (when talking about the requirements of the task)		MA
Strategy formulation	Specific discussion about the details of the content of the task		SF
Monitoring progress towards goals	Monitoring tasks and the process and communicating progress to team members	Explicit communication: intentional information sharing among team members (task status – time, done)	MPEC
		Verbal shadowing (narration): when someone narrates what they are doing or what they are going to do	MPVS
System monitoring	Monitoring of events in the work area.	Basic awareness (awareness:) keeping an eye on who is present in the work area, what they are doing and where they are doing it – who sees what, who hears what	SMBA
		Consequential communication: information communicated by objects in the workspace – information communicated, noticed, received by objects or avatars: e.g., something happened to the text (it appeared, disappeared, is not visible) or changed its colour, location, etc.	SMCC
		Other: e.g., where a particular function is, feedback to the system, whether the system notifies about something, or whether someone has accidentally moved or deleted something	SMO

Team monitoring and back-up behaviour	Helping each other, on request or without request		TM
Coordination	Coordination of actions: what you or us should do (imperative mode)		C
Conflict management	Conflict management: apologising or joking away		CM
Motivation and confidence building	Building shared confidence (praising each other)		MB
Affect management	Regulation of members' emotions		AM
Confirmation	Confirmation		CO
Researcher	When the researcher speaks and how the participants react to it		R
No data	No data		ND

Table 2. The code system used for the analysis of communication transcripts (created by the authors)

Due to the marginal nature of MA, we merged it with another code, SF (strategy formulation). According to the Marks, Mathieu and Zaccaro (2001) framework, both codes (MA and SF) fall into the transition category and are therefore considered similar codes. On this basis, the change in the reliability of the code system is presented in Table 3 ($\kappa = 0.823$; $p < 0.001$). The kappa value increased minimally from 0.819 to 0.823.

Code abbreviation	ICC value	p	Code reliability
AM	0.920	$p < 0.001$	excellent
MPEC	0.667	$p < 0.001$	medium
MPVS	0.841	$p < 0.001$	good
C	0.748	$p < 0.001$	medium
CM	0.894	$p < 0.001$	good
R	0.961	$p < 0.001$	excellent
CO	0.837	$p < 0.001$	good
MB	0.882	$p < 0.001$	good

ND	1.000	p<0.001	excellent
SMBA	0.706	p<0.001	medium
SMO	0.807	p<0.001	good
SMCC	0.703	p<0.001	medium
BB	0.840	p<0.001	good
SF	0.799	p<0.001	good

Table 3. Reliability of the code system used for the analysis of the communication transcripts after merging the MA and SF codes (created by the authors)

As demonstrated in Table 3, we used a total of 14 codes for the sequence analysis. Based on the original 14-code coding scheme, no significant results were found in 2 out of 7 teams in the first lab study and in 6 out of 11 teams in the second lab study. This could indicate several things. On the one hand, it may not be surprising that there are no communication patterns since we cannot talk about structured work in the traditional sense. Communication patterns may be influenced by familiarity, task and instrument. However, the coding system also influences the results. With such a fine-tuned code system of 14 elements, significant communication patterns cannot be detected in 8 out of 18 teams. For this reason, we decided to transform the original code system of 14 codes into a code system of 10 codes, as described below. In the transformation, the frequency of each code, its reliability and whether it appeared in all teams were also taken into account.

Thus, a new 10-element code system was created (Table 4), which also has high reliability ($\kappa = 0.825$; $p < 0.001$).

Code abbreviation	Code name	ICC value	p	Code reliability
AM	Affect management	0.921	p<0.001	excellent
PM	Progress monitoring	0.804	p<0.001	good
C	Coordination	0.748	p<0.001	medium
R	Researcher	0.961	p<0.001	excellent
CO	Confirmation	0.837	p<0.001	good
MB	Motivation and confidence building	0.882	p<0.001	good
SMBA	System monitoring basic awareness	0.718	p<0.001	medium
SMO	System monitoring other	0.807	p<0.001	good
TM	Team monitoring and assistance	0.840	p<0.001	good
SF	Strategy formulation	0.799	p<0.001	good

Table 4. Code system used for sequence analysis (created by the authors)

4.2. Emergence of team communication patterns in teams participating in the first laboratory study

In the sequence analysis, our aim was to see if the communication patterns of the teams differ when certain types of usability problem arise. The main difference between teams is whether or not they experienced the overwriting problem. We consider overwriting to be the most serious team usability problem, so we divided the teams into two groups according to whether overwriting appeared during their work or not. Overwriting is also special because this type of problem appears only in the communication transcripts.

The results of the sequential analysis confirm our hypothesis, that the teams where overwriting occurred have different communication patterns.

We used the Yule Q value, ‘which is the recommended method for statistically quantifying sequential associations between two events’ (Lloyd, Kennedy and Yoder 2013, 480). The Yule Q value can take a value between -1 and 1, indicating the strength and direction of the relationship between the two codes. A -Q value indicates a negative relationship between the codes, while a +Q value indicates a positive relationship.

Overwriting was reported as a problem in Teams 1, 2, 3 and 5 in the first laboratory study (Table 5).

Based on their communication patterns, the teams that did not experience overwriting (Teams 4 and 6) are those in which:

- the SMBA (system monitoring basic awareness) code is followed in almost all cases (above 0.9, i.e., a very strong Yule Q value) by another SMBA code, meaning that if someone communicates or requests information about where they are or what they are doing in the workspace, their peers respond with this type of information, so there is a discourse among the participants about it, which alone is enough to avoid a conflict;
- the TM (help) and C (coordination) codes also play a significant role – overwriting can be avoided where the SMBA–SMBA value is strong (between 0.5 and 0.69), but not above 0.9, and the TM and C codes are combined (!).

	Team 1	Team 2	Team 3	Team 4	Team 5	Team 6	Team 7
Overwriting	✓	✓	✓		✓		
Significant patterns detected (p-value)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SMBA–SMBA	0.628	0.792	0.762	0.943	0.937		0.873
TM–TM	0.76		0.382	0.714	0.789	0.707	0.624
C–C							0.844

Table 5. Team communication patterns for each team – Yule Q values of the first lab study (created by the authors)

Team 5 is an exceptional team as overwriting occurred here despite its having an SMBA–SMBA above 0.9 and a high TM–TM. This can be explained by the fact that this team was the only one to have a synchronisation problem and a very pronounced problem with verbal communication (Skype was interrupted). For this reason, despite all efforts, overwriting appeared.

4.3. Emergence of team communication patterns in the teams participating in the second laboratory study

In the second laboratory study, Teams 1, 2, 4, 5, 6 and 7 experienced the overwriting problem, while Teams 3, 8, 9, 10 and 11 did not (Table 6). The result of the first lab study is thus strengthened: teams that did not experience the overwriting problem can be characterised by the same communication patterns, that is, there was a discourse about awareness, or team members helped each other effectively, or they organised and planned collaborative work closely.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
Overwriting	✓	✓		✓	✓	✓	✓				
Significant patterns detected (p-value)	<0.01	<0.01	<0.01		<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
SMBA–SMBA	0.694	0.895	0.938		0.604	0.804	0.844	0.937	0.905	0.609	0.916
TM–TM	0.378		0.315			0.447			0.605	0.555	0.882
C–C		0.364								0.675 (C–CO)	

Table 6. Team communication patterns for each team – Yule Q values of the second lab study (created by the authors)

Sequence analysis can be used to better understand and investigate team usability problems. It can reveal the communication patterns behind the problems, thus providing relevant results about different teams and team usability problems. Sequence analysis is therefore considered an important part of Team Usability Testing.

5. Discussion

Although team communication is a significant topic of study in the psychological literature, it is a less researched area in relation to software usability. In this paper, we investigated team communication patterns in the context of software usability: we

performed communication analysis of teams using shared online visual workspaces. We used sequence analysis to explore the communication patterns of the teams using PREZI or Miro.

An analysis of the communication patterns related to all of the usability problems is beyond the scope of this article, so we selected the overwriting problem. Our choice is justified by the fact that this is the most severe team usability problem; it occurs when one participant inadvertently (due to inadequate support of collaborative features) overwrites or deletes the work of another participant in the collaborative visual workspace. Overwriting was avoided by some teams but not by others, so we examined the nature of the communication patterns between the teams where overwriting occurred and the teams where it did not.

Our results show that for those teams where there is a discourse on situational awareness, the overwriting problem does not appear. So, if someone communicates or requests information about what is happening in the shared visual workspace, team members respond with this type of information. In addition, teams that are effective in helping each other (if a team member asks for help, s/he gets help) or have a tight organisation and planning of collaborative work (if a team member shares information about the organisation of collaborative work, s/he also receives this type of information in response) can avoid overwriting.

The results are in line with previous research results, which claim that by identifying communication patterns, we can reveal the communication dynamics of each group and thus identify difficulties, stagnations and problems (Juhász 2015; Hámornik 2013; Soós 2012).

The novelty of the results is that we have examined this in the context of software usability. Just as certain conflicts and problems occur in teams that communicate differently, this is also true for the usability testing situation: teams with certain communication patterns will experience different usability problems. It can also have an impact on the interpretation of the usability test results, as teams' communication strategies can obscure serious usability problems.

Team communication in the context of software usability is a poorly researched topic, so during the research we developed a suitable code system by combining team process theory and mechanics of collaboration theory (Marks, Mathieu and Zaccaro 2001; Pinelle, Gutwin and Greenberg 2003). By merging the two code systems, we developed our own code system, which is described in detail in the Results section. The reliability of the final code system is high (Cohen's kappa = 0.825; $p < 0.001$) (Table 4), making it suitable for exploring communication patterns of teams in the context of software usability. In summary, the new code system can reliably identify team communication patterns in the context of usability and is therefore considered part of the Team Usability Testing method.

6. Limitations and further research

The type of software evaluated in our study may limit the scope of our conclusions. Thus, as a next step of our research, it would be important to expand the number

and types of collaborative software evaluated. Examining multiple types of software would make it easier to distinguish between usability problems specific to the software under investigation and general problems common to similar types of collaborative software. At the same time, it would also be possible to investigate the coding system while evaluating other types of collaborative software, broadening the scope of applicability.

References

- Adams, Shawnette K. 'Disciplinarily Hetero- and Homogeneous Design Team Convergence: Communication Patterns and Perceptions of Teamwork'. Doctoral Dissertation, Virginia Tech, 2007.
- Aldag, Ray, and Loren Kuzuhara. *Creating High Performance Teams: Applied Strategies and Tools for Managers and Team Members*. New York: Routledge, 2015.
- Bafoutsou, Georgia, and Gregoris Mentzas. 'Review and Functional Classification of Collaborative Systems'. *International Journal of Information Management* 14, no. 4 (2002): 281–305.
[https://doi.org/10.1016/S0268-4012\(02\)00013-0](https://doi.org/10.1016/S0268-4012(02)00013-0)
- Cohen, Susan G, and Diane E Bailey. 'What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite'. *Journal of Management* 23, no. 3 (1997): 239–90.
- Cooke, Nancy J., Andrew Duchon, Jamie C. Gorman, Joann Keyton, and Anne Miller. 'Preface to the Special Section on Methods for the Analysis of Communication'. *Human Factors* 54, no. 4 (2012): 485–8.
<https://doi.org/10.1177/0018720812448673>
- Fehér, Katalin. 'Okos Város: Trendtémák És Konceptiók'. *Információs Társadalom* 17, no. 4 (2017): 25–38.
- Geszten, Dalma, Balázs Péter Hámornik, and Károly Hercegfí. 'Exploring Awareness Related Usability Problems of Collaborative Software with a Team Usability Testing Approach'. In *Proceedings of the 9th IEEE International Conference on Cognitive Infocommunications, CogInfoCom*, 45–50. Budapest, Hungary: IEEE, 2018.
<https://doi.org/10.1109/CogInfoCom.2018.8639865>
- Geszten, Dalma, Balázs Péter Hámornik, and Károly Hercegfí. 'Usability Evaluation of a Collaborative Design Software in the Wild'. In *Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications, CogInfoCom*, 101–6. Naples, Italy: IEEE, 2019.
- Geszten, Dalma, Balázs Péter Hámornik, and Károly Hercegfí. 'Empirical Study of Team Usability Testing: A Laboratory Experiment'. *Cognition, Technology & Work* 23, no. 4 (2021): 1–15.
- Geszten, Dalma, Balázs Péter Hámornik, and Károly Hercegfí. 'Finding Treasure in the Wild: A Field Study of a Collaborative Design Software'. In *Accentuated Innovations in Cognitive Info-Communication. Topics in Intelligent Engineering and Informatics*, edited by Ryszard Klempous, Jan Nikodem, and Péter Zoltán Baranyi, 185–202. Cham: Springer, 2023.
- Hackman, J Richard. 'The Design of Work Teams'. In *Handbook of Organizational Behavior*, edited by Jay William Lorsch, 315–42. Englewood Cliffs, NJ: Prentice-Hall, 1987.

-
- Hámornik, Balázs Péter. 'Team Tudás, Interakció És Kommunikáció Vizsgálata: Tudásmegosztás Az Orvosi Rehabilitációs Teamben'. Doctoral Dissertation, ELTE, 2013.
- Hámornik, Balázs Péter, Máté Köles, Anita Komlódi, Károly Hercegfí, and Lajos Izsó. 'Features of Collaboration in the VirCA Immersive 3D Environment'. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics*, 130–9. Krakow, Poland: AHFE, 2014.
<http://www.ahfe2014.org/files/books/2014CN.pdf>
- Hercegfí, Károly, Anita Komlódi, Bálint Szabó, Máté Köles, Emma Lógó, Balázs P. Hámornik, and Gyöngyi Rózsa. 2015. 'Experiences of Virtual Desktop Collaboration Experiments'. In *Proceedings of the 6th IEEE International Conference on Cognitive Infocommunications, CogInfoCom*, 375–9. Győr, Hungary: IEEE, 2015.
<https://doi.org/10.1109/CogInfoCom.2015.7390622>
- Hertel, Guido, Susanne Geister, and Udo Konrad. 'Managing Virtual Teams: A Review of Current Empirical Research'. *Human Resource Management Review* 15, no. 1 (2005): 69–95.
<https://doi.org/10.1016/j.hrmr.2005.01.002>
- Hewes, Dean E, Marshall Scott Poole, and A B Hollingshead. 'The Analysis of Group Interaction Processes'. In *Research Methods for Studying Groups and Teams: A Guide to Approaches, Tools, and Technologies*, edited by Andrea Hollingshead and Marshall Scott Poole, 358–85. New York: Routledge, 2012.
- Juhász, Márta. 'Kogníció És Kommunikáció a Teammunkában'. *Alkalmazott Pszichológia* 15, no. 3 (2015): 5–27.
<https://doi.org/10.17627/ALKPSZICH.2015.3.5>
- Keszei, Barbara, Zsófia Böddi, Dalma Geszten, Balázs Péter Hámornik, and Andrea Düll. 'A Krippendorff-Alfa (KALPHA) Alkalmazása a Gyakorlatban'. *Alkalmazott Pszichológia* 19, no. 2 (2019): 57–75.
- Khoshafian, Setrag, and Marek Buckiewicz. *Introduction to Groupware, Workflow, and Workgroup Computing*. Hoboken, NJ: John Wiley & Sons, 1995.
- Kirkman, Bradley L, and John E Mathieu. 'The Dimensions and Antecedents of Team Virtuality'. *Journal of Management* 31, no. 5 (2005): 700–18.
- Krumm, Stefan, Jens Kanthak, Kai Hartmann, and Guido Hertel. 'What Does It Take to Be a Virtual Team Player? The Knowledge, Skills, Abilities, and Other Characteristics Required in Virtual Teams'. *Human Performance* 29, no. 2 (2016): 123–42.
<https://doi.org/10.1080/08959285.2016.1154061>
- Laufer, László, Péter Halácsy, and Ádám Somlai-Fischer. 'Prezi Meeting: Collaboration in a Zoomable Canvas Based Environment'. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 749–52. Vancouver, Canada: CHI EA '11, 2011.
<https://doi.org/10.1145/1979742.1979673>
- Levi, Daniel, and David A Askay. *Group Dynamics for Teams*. London: Sage Publications, 2020.
- Lloyd, Blair P., Craig H. Kennedy, and Paul J. Yoder. 'Quantifying Contingent Relations from Direct Observation Data: Transitional Probability Comparisons versus Yule's Q'. *Journal of Applied Behavior Analysis* 46, no. 2 (2013): 479–97.
<https://doi.org/10.1002/jaba.45>
- Marks, Michelle A, John E Mathieu, and Stephen J Zaccaro. 'A Temporally Based Framework and Taxonomy of Team Processes'. *Academy of Management Review* 26, no. 3 (2001): 356–76.

- Marlow, Shannon L., Christina N. Lacerenza, and Eduardo Salas. 'Communication in Virtual Teams: A Conceptual Framework and Research Agenda'. *Human Resource Management Review* 27, no. 4 (2017): 575–89.
<https://doi.org/10.1016/j.hrmr.2016.12.005>
- Martins, Luis L., Lucy L. Gilson, and M. Travis Maynard. 'Virtual Teams: What Do We Know and Where Do We Go from Here?'. *Journal of Management* 30, no. 6 (2004): 805–35.
<https://doi.org/10.1016/j.jm.2004.05.002>
- Mesmer-Magnus, Jessica R, Leslie A DeChurch, Miliani Jimenez-Rodriguez, Jessica Wildman, and Marissa Shuffler. 'A Meta-analytic Investigation of Virtuality and Information Sharing in Teams'. *Organizational Behavior and Human Decision Processes* 115, no. 2 (2011): 214–25.
- O'Connor, Brian P. 'Simple and Flexible SAS and SPSS Programs for Analyzing Lag-Sequential Categorical Data'. *Behavior Research Methods, Instruments, & Computers* 31, no. 4 (1999): 718–26.
- Pinelle, David, and Carl Gutwin. 'Groupware Walkthrough'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Changing Our World, Changing Ourselves*, 455–62. Minneapolis, MN, USA: CHI '02, 2002.
<https://doi.org/10.1145/503376.503458>
- Pinelle, David, Carl Gutwin, and Saul Greenberg. 'Task Analysis for Groupware Usability Evaluation'. *ACM Transactions on Computer–Human Interaction* 10, no. 4 (2003): 281–311.
<https://doi.org/10.1145/966930.966932>
- Salas, Eduardo, C Shawn Burke, and Janis A Cannon-Bowers. 'Teamwork: Emerging Principles'. *International Journal of Management Reviews* 2, no. 4 (2000): 339–56.
- Salomón, Sergio, Rafael Duque, Jose Luis Montaña, and Luis Tenés. 'Modeling Users Behavior in Groupware Applications'. In *Cooperative Design, Visualization, and Engineering. CDVE 2019. Lecture Notes in Computer Science*, edited by Yuhua Luo, 11–21. Cham: Springer, 2019.
https://doi.org/10.1007/978-3-030-30949-7_2
- Saunders, Carol S. 'Virtual Teams: Piecing Together the Puzzle'. In *Framing the Domain of IT Management: Projecting the Future through the Past*, edited by Robert W. Zmud, 29–50. Cincinnati, OH: Pinnaflex, 2000.
- Schmidt, Kjeld, and Liam Bannon. 'Constructing CSCW: The First Quarter Century'. *Computer Supported Cooperative Work: CSCW: An International Journal* 22, no. 4–6 (2013): 345–72.
<https://doi.org/10.1007/s10606-013-9193-7>
- Szoós, Juliánna Katalin. 'Team – Folyamatok Elemzése Operátor Teameknél: A Team-Folyamatok Vizsgálata Magas Kockázatú Szervezetekben Dolgozó Operátor Teameknél'. Doctoral Dissertation, ELTE, 2012.
- Steves, Michelle Potts, Emile Morse, Carl Gutwin, and Saul Greenberg. 'A Comparison of Usage Evaluation and Inspection Methods for Assessing Groupware Usability'. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, 125–34. Boulder, CO, USA: GROUP '01, 2001.
<https://doi.org/10.1145/500286.500306>
- Szűts, Zoltán, and Jinil Yoo. 'A Chatbotok Jelensége, Taxonómiája, Felhasználási Területei, Erősségei És Kihívásai'. *Információs Társadalom* 18, no. 2 (2018): 41–55.
<https://doi.org/10.22503/inftars.xviii.2018.2.3>

Tófalvy, Tamás. ‘„Sokkal Több Is Lehetett Volna, Mint Egy Óriási iPhone’. *Információs Társadalom* 19, no. 1 (2019): 92–107.

<https://doi.org/10.22503/inftars.xix.2019.1.4>

Vajda, Kinga. ‘Az Idős Generáció És Az Infokommunikáció Kapcsolata Az Idősek Nappali Ellátásában – Szociális Szolgáltatásfejlesztési És Prevenációs Perspektívák Egy Hazai Kutatás Tükrében’. *Információs Társadalom* 20, no. 3 (2020): 71–91.

Yen, David C., H. Joseph Wen, Binshan Lin, and David C. Chou. ‘Groupware: A Strategic Analysis and Implementation’. *Industrial Management and Data Systems* 99, no. 1 (1999): 64–70.

<https://doi.org/10.1108/02635579910243879>

Hume's Empiricism versus Kant's Critical Philosophy (in the Times of Artificial Intelligence and the Attention Economy)

The article exposes how one of the most fundamental oppositions in the history of modern philosophy – the opposition between Hume's empiricism and Kant's critical philosophy – is embedded in the current debate on the impact of artificial intelligence (in particular, the algorithmic selection of content) on human society. Hume's empiricism – with its deduction of subjectivity based on a process of habituation – corresponds to the functioning of recommending algorithms, while Kant's idea of autonomous subjectivity corresponds to the ideals underlying today's ethical attempts towards the regulation of artificial intelligence. According to such ethics, the use of empirical data can endanger humans; whereby our attention can be easily caught by sensationalist content and our autonomy replaced by the agency of machinic intelligence. However, as argued in the present article, such ethical positioning also reproduces the gap between the empirical reality and normative principles, which is why transcendental (Kantian) ethics should be supplemented with Hume's immanent and practical reasoning.

Keywords: *Habit, Recommendation Algorithms, Attention Economy, Artificial Intelligence, David Hume, Immanuel Kant*

Acknowledgment

This work received financial support from the Slovenian Research Agency – ARRS (Research program *Mass media, public sphere and social changes*, grant No. P5-0051 (B)).

Author Information

Jernej Kaluža, Social Communication Research Centre, Faculty of Social Sciences, University of Ljubljana, Slovenia; Jernej.Kaluza@fdv.uni-lj.si;
<https://orcid.org/0000-0003-1781-3713>

How to cite this article:

Kaluža, Jernej. "Hume's Empiricism versus Kant's Critical Philosophy (in the Times of Artificial Intelligence and the Attention Economy)".

Információs Társadalom XXIII, no. 2 (2023): 67–82.

==== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.4> ====

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

The “empirical turn” in the philosophy of technology, contrary to the opinion of some (see Cera 2020), may not imply the abandonment of all the great ontological questions from the history of philosophy. On the contrary, as a review of the vast literature that has followed the rise in the use of artificial intelligence in recent years reveals, it is possible to find deep philosophical problems embedded even in quite concrete technological innovations. An excellent illustrative example of a concrete new technology, the emergence of which has raised several major philosophical questions and ethical dilemmas, may be “AI-powered content recommendation engines” (see, among others, Sorbán 2021; Carlson 2018; Chun 2016; Kant 2020). Since the first studies and popular books on the subject, the algorithmic personalisation of content has sparked heated philosophical debates on various issues, particularly those related to the conceptions of privacy and publicity, subjective autonomy, and social power. Initially, the debate was mainly framed around issues of algorithmic surveillance, the right to privacy, and the potential misuse of the data obtained, which led to a re-actualisation of Foucault’s thesis on the panoptic control of society. In parallel, the rise of algorithmisation and datafication has been accompanied by debates on cognitive biases, technological addiction, and political radicalisation (see for example, Williams 2018; Tufekci 2015).

However, the topicality and acuteness of the developments involving algorithms and the use of artificial intelligence have meant that attention has been repeatedly drawn to various empirical incidents that have arisen associated with their use (for example, fake news, hate speech, and political polarisation), while systematic and deeper attempts to philosophically ground the issues have been rarer. In this article, we aim to do just that: to highlight how one of the most fundamental oppositions in the history of modern philosophy, namely the opposition between Hume’s empiricism and Kant’s critical philosophy, is deeply embedded in today’s debates on the impact of artificial intelligence (and in particular the algorithmic personalisation of content) on human society. As further explicated below, Hume’s empiricism, with its deduction of subjectivity to the predictability of habit, i.e. to past experiences from which the patterns of future action can be predicted, can be seen to correspond to the big data-based operation of recommendation algorithms. On the other hand, Kant’s idea of autonomous, rational, and free subjectivity corresponds to the ideals underlying today’s ethical attempts to regulate and constrain the functioning of artificial intelligence.

Focusing on Hume’s concept of habit seems appropriate for conceptualisation of the complex interaction between human agency and artificial intelligence. It also corresponds with the focus on other pertinent issues, such as the attention economy, cognitive biases, and mental health (which seem to be replacing the exclusive focus on surveillance as the main political problem related to recommending algorithms in the last few years). As Tanya Kant (2020, 9) argued: “Data tracking does not exist, in and of itself, simply to surveil or track users, but to anticipate them /.../, to ‘know’ some facet(s) of a user’s identity in order to make ‘personally’ relevant some component of experience on their behalf” (see also Cheney-Lippold 2017). In this

text, I argue that the concept of habit, with all its associated contradictions, could provide a fundamental philosophical grounding, from the perspective of which one can comprehend the broader complexity of the problems associated with algorithmic content distribution (see also Kaluža 2022a; 2022b). Additionally, the concept of habit opens up questions related to the principles of ethical behaviour in times of the attention economy.

This article, based on a conceptual analysis of texts from the history of continental philosophy and contemporary publications in the scientific fields of media and communication studies, is structured as follows: The first part describes how the concept of habit is explicitly or implicitly situated in contemporary critical theory that addresses the problematic effects of algorithmic personalisation. The second part briefly situates the concept of habit within Hume's empiricism and shows how subjectivity can be influenced through habituation (as these ways are also used by contemporary, big data-based predictive analysis of human behaviour). The third part of the article explicates how Kant's response to Hume's philosophy, which seeks to establish a sphere of purity independent of habit, can be understood in a way analogous to the response to a certain part of the ethics debate today, which urges users to confront algorithms with the power of their free will. This kind of ethical positioning, as I argue in the concluding section of the paper, often reproduces the gap between empirical reality and normative principles, which is why I advocate that transcendental (Kantian) ethics should be supplemented with critical empiricism, as a form of immanent ethics derived from concrete empirical situations and which aims to change the state of affairs "from within".

2. The concept of habit in the debate on the algorithmic distribution of content

Algorithmic content selection challenges human agency and is increasingly replacing human curators of content. "AI-powered recommendation systems" have become an increasingly key focus of interest as AI has "recently reached the level of development that makes their functioning comparable to human thinking and allows them to perform tasks requiring (close to) human intelligence" (Sorbán 2021). Such assumptions, however, have not only remained at an abstract level, but have to a certain extent also materialised in a very concrete way in the modern world, as algorithmic content selection is now replacing some of the professions that are closest to the ideals of human curation and decision-making, such as journalists and editors (Carlson 2018; Vobič, Šikonja, and Kalin Golob 2019). Artificial intelligence "which is getting more and more independent from humans through machine learning methods", could, in this context, be considered "a major threat to the human race" as seen in "countless science fiction movies in which AI machines try to take control over mankind" (Hosseinpour 2020, 49–50).

However, the paradox is that this struggle between AI and humans is not caused by the disobedience of AI. On the contrary, it is fuelled by treating AI as a purely subordinate "tool or servant" (see Gyulai in Ujlaki 2021; Totschnig 2019). The superi-

ority of AI for content selection is conditioned precisely by the ability of technology to please the interests of users, while human agency is passivised (individual users) or excluded from the process altogether (professional human content curators, such as journalists). In this trickery, we can see the manifestation of “the tragedy of the master” (Coeckelbergh 2015), where, based on Hegel’s dialectic between master and slave, all the work is left to the slave, while the master is completely passivised and becomes alienated from his or her own mastering agency. One could also argue that we are now faced with the realisation of Heidegger’s reasoning that human thought – in its own calculating striving for domination – is becoming more and more subordinated to technology (Hosseinpour 2020).

However to date, less attention has been paid to the fact that this subjugation of humans to technology is today primarily conditioned by the habituation of the subject to technology made possible by artificial intelligence (machine learning, neural networks, etc.). Recommendation algorithms are a particularly good illustrative example in this context since they work by exploiting users’ previous habits (recorded in the form of data on their history of online activity) to serve the users and recommend future activity. Such service also implies a passivisation of the user. It is this passivisation through repetitive activity, which leads to rigid and predetermined (one might even say programmed) behaviour, that has been one of the key features of the concept of habit that has been so discussed in the history of philosophy (see Malabou 2004; Chun 2016).

The concept of habit is also implicitly present in the debates on various cognitive biases, which are supposed to be encouraged by the algorithmic selection of content. One of the most well-known of these is the so-called confirmation bias, namely, the tendency to confirm one’s past attitudes, beliefs, and values, which occurs mainly in the processes of searching for, filtering, and interpreting information. For example, Ariely (2008, xx) showed that there is a certain rigidity not only in the rational, but also in irrational thought, stating: “irrational behaviors of ours are neither random nor senseless. They are systematic, and since we repeat them again and again, predictable”. This kind of thesis implies that there must be an instance – and habit seems to be an appropriate concept to name it – that makes the opposite of rational thought not chaotic and unpredictable, but predictably irrational. Habitual decision is the opposite of conscious, thoughtful, and rational decision: It follows certain patterns and can thus also be manipulated. This persuasive manipulation, which could be understood as a “hidden influence that attempts to interface with people’s decision-making processes in order to steer them toward the manipulator’s ends” does not “persuade people” rationally, but rather seeks to exploit their irrational “vulnerabilities”, such as “hidden fears” (Lewandowsky et al. 2020, 23). Seaver (2018), in his analysis, went even further and claimed that the major tendency among today’s developers of recommender systems is to “hook” people: He considers recommender systems as traps and compares their functioning with animal trapping based on anthropological theories.

The philosophical dichotomy between rationality and irrationality and the hierarchy between different levels of self-consciousness and self-possession are therefore implicitly ever-present in discussions on algorithmic personalisation. Even if

the creators of technologies “often justify their design decisions by saying they’re ‘giving users what they want’, critical theory doubts that they really give users ‘what they want to want’” (Williams 122). With such an argument, the tension between users’ “higher” rational goals and users’ “lower” irrational passions, which often go against the individual’s interest, is confirmed. And it is this tension, which is indirectly connected with Kant’s rejection of Hume’s reduction of higher instances of mind to habitual functioning (see Deleuze 1967/2008), that is today most directly manifested in clickbait logic or in the logic of nudging, i.e. “influencing action by subtle prompts and signals” (Couldry and Mejias 2019, 140). It is not a coincidence, therefore, that behavioural economics is becoming a crucial part of the broader marketing profession. This development, though, has led to warnings that the potential “integration of neuroscience and marketing” could bypass “the speaking human subject” (Couldry and Mejias 2019, 141).

Kant’s autonomy of the human mind has perhaps never been so severely challenged throughout history, since the struggle for attention requires the achievement of ends regardless of the means: “All these types of distractions undermine people’s autonomy, by instilling habits and desires that are not voluntarily chosen. In other words, persuasive design steers users’ attention towards irrational behaviours” (Voinea et al. 2020, 2354). This division between the higher (reason, rational thought) and lower (habit, instinct, desire) instances of the mind is in contemporary debate often expressed through the vocabulary of the “attention economy”, according to which our attention is alienated from us through the use of psychological tricks that count on “our capacity to fool ourselves” (Galef 2021, 7).

The balance between a more goal-oriented reality principle on the one hand, and pleasure principle which aims for the immediate fulfilment of desire on the other hand, seems to be disturbed by our failure to fully distinguish rationality from desire and emotional from intellectual activity. However, this inability can not only influence our individual attention, but it can have crucial effects on broader society too. For example, it has radically changed the news industry by encouraging business models that are based on the monetisation of attention (see Myllylahti 2020; and Slaček Brlek 2018) and it has even led to a broader phenomenon related to the quantification of the public sphere (Splichal 2022). Data behaviourism, which is replacing “disinterested social knowledge”, is also affecting the broader sphere of social sciences, in which “social knowledge becomes whatever works to enable private or public action to modulate others’ behavior in their own interests” (Couldry and Mejias 2019, 128).

As a response to this development, today’s critics of the algorithmically driven dominance of the instantaneous, the shocking, and the clickable, argue for the promotion of the subject’s autonomy, morality, and rationality, but understood in a specifically Kantian way. By this we mean (as it will be further developed) Kant’s fundamental philosophical orientation towards the establishment of a purified sphere that is separated and independent from anything empirical. On the other hand, the understanding and the design of artificial intelligence follows the logic of a distinctly Humean notion of the mind. Hume’s empiricism, in which human subjectivity is formed solely through experience, corresponds with the basic idea

in data harvesting, machine learning, and pattern recognition techniques: The idea that habitual repetition in the past indicates a tendency towards repetition of the same also in the future.

3. Hume's concept of subjectivity and the big-data turn

It seems pertinent to ask: What can philosophical thought from almost 300 years ago and current trends in data science and artificial intelligence development possibly have in common? What is behind Wendy Chun's strange thesis that "Hume is the favored philosopher of Big Data analytics" (Chun 2016, 54)? And why did Hume's thoughts, as they took shape in his works *A Treatise of Human Nature* (1739/2000) and *An Enquiry concerning Human Understanding* 1748/2004), provoke similar moral and epistemological concerns that the rise of algorithms and big data are provoking today?

As a starting point from which answers to these questions will be developed, we could consider the radical nature of Hume's empiricism, which grounds "human nature entirely upon experience" (Hume 1739/2000, 407) and proposes that all our ideas derive from our impressions: "We can never think of anything which we have not seen without us, or felt in our minds" (ibid. 408). To put the same point in a different way: All that we know (or that we believe that we know) is derived from what we have encountered along our past path. "It follows, then, that all reasoning concerning cause and effect are founded on experience, and that all reasonings from experience are founded on the supposition, that the course of nature will continue uniformly the same" (ibid. 410).

That is to say, our current beliefs, expectations, and customs are what they are, therefore, solely because of our previous experiences. This is why Hume's empiricism reveals the fundamental "emptiness of the self" (Chun 2016, 6). As Deleuze (1977/1991, x) wrote in his book on Hume: "We are habits, nothing but the habit of saying 'I'". Subjectivity is thus formed through experience, and this implies that users with similar experiences – for example, those who have a similar search history, or who like similar things or live in a certain geographical area – can be compared. As many authors have mentioned also in the context of today's data turn, "your profile is only made meaningful and commodifiable to marketers in and alongside the context of other users' profiles" (Kant 2020, 34–35). It is this kind of "plasticity of subjectivity" (Malabou 2004), which implies that users can be changed by orienting their experience and manipulating their habits, that has guided the conception of many Silicon Valley products (see Eyal 2014).

Subjectivity as understood by Hume is thus predictable: We believe that something will repeat in the future because we are accustomed or habituated to the repetition of the same in the past. Custom or habit is, according to Hume (1739/2000, 411), "the guide of life". It "determines the mind, in all instances, to suppose the future conformable to the past" (ibid. 411). This same proposition presents the grounding of the algorithmisation and big data turn of today; whereby data representing past experiences functions as a basis for a recommendation of the same (or similar) con-

tent in the future. “Algorithmic personalization is premised on the idea that your future preference scan be inferred from your past interactions” (Kant 2020, 37) and that a user’s previous online activity is the best predictor of what he or she wants to see next.

Hume’s empiricism, in which human subjectivity is formed solely through experience, corresponds with the basic idea behind data harvesting, machine learning, and pattern recognition techniques. Hume’s subjectivity is reduced to experience and online subjectivity of today is reduced to data (Cheney-Lippold 2017). This is the reason why data analytics are, according to Chun (2016, 57), usually related to habit: They focus on habitual actions (such as buying), they seek to change habits, and they “replace” causality with correlations between habits. This “process of commercial anticipation” was somehow neglected in the early problematisation of algorithmisation, which focused mostly on the issue of surveillance: the central aim of data harvesting is not just to watch an individual users, “but instead to act on [them], with, or against their experience of the web” (Kant 2020, 9).

4. Kant’s answer to Hume’s provocation: the autonomy and purity of subjectivity

It is exactly the predictability of subjectivity, which derives from its reductibility to previous experience, that was one of the main provocations of Hume’s philosophical gesture as represented in the context of the history of philosophy. Kant’s critical project could thus be understood as a conceptual counterpart to Hume’s empiricism. As Kant admits, Hume interrupted his “dogmatic slumber” and gave completely different directions to his “researches in the field of speculative philosophy” (Kant 1783/2004, 10). However, to avoid Hume’s sceptical conclusions, Kant argued that some concepts do not derive from experience, but from “pure understanding” (Ibid. 10). In general, Kant’s “new science”, for which Hume gave a “hint” (even if he latter “deposited his ship on the beach [of skepticism]”) (ibid. 11–12), aims to re-establish the autonomy of human understanding, reason, and moral judgement.

Kant’s project could therefore be understood as a continuation of Hume’s empiricism (since it rejects old metaphysical notions of cause, God, and essence), but also as a negation of the empiricist understanding of human nature, which deprives human agency of autonomy and self-determination (see Deleuze 1967/2008). Our behaviours and beliefs namely seem in Hume’s world to be defined by experience, by our encounters with the outside world, while Kant re-establishes the sphere of “purity” (the sphere of pure reason, pure judgement, pure understanding), which is defined exactly by its independence from anything empirical. As he wrote in the preface to the first edition of the *Critique of pure reason*, his project presents a critique “of the faculty of reason in general, in respect of all knowledge after which it may strive independently of all experience” (Kant 1781/1996, 8).

It is difficult to pin down exactly why Hume’s thoughts were perceived as a scandal, since there is nothing particularly provocative or transgressive in his style, as there is, for example, with Nietzsche or Marx. Nevertheless, the reasons for this

can be found in a specific combination of Hume's implicit scepticism, determinism, and amoralism. Scepticism stems from Hume's rejection of the idea of a necessary connection between cause and effect (Hume 1748/2004, 42–43). This rejection is also manifested in today's data turn, which is based on the principle of correlation instead of on the principle of causality (see Splichal 2022; and Mayer-Schönberger and Cukier 2013). In Hume's world, we have no direct knowledge of the interconnectedness of phenomena. Similarly, one cannot speculate about the causes for patterns that are observed in big data, but can only draw correlations between them. "Unlike earlier modes of governance", David Chandler (2019, 29) argued, "digital governance does not seek to make causal claims" and it "increasingly focus[es] on the 'what is' of the world in its complex and plural emergence".

The consequence of such a functioning of social knowledge, which is "built from unstructured data, drawn directly from the traces left in the flow of everyday life" (Couldry and Mejias 2019, 125) is specific scepticism; whereby "the end product (the knowledge generated) is not explicable in terms of rules /.../ no one – not even the engineers who run the process – can account exactly for how that knowledge was generated" (Ibid.). Hume's abandonment of the necessary link between cause and effect is, besides scepticism, linked also with his associationism, according to which the main ability of the mind is to habitually associate ideas. The sceptical character of such reasoning is derived from its inability to point to a first cause, which could transcend experience. Even Hume's principle of habit as the central concept of his thoughts in this area should not be understood in such a fashion: "By employing that word, we pretend not to have given the ultimate reason of such a propensity. We only point out a principle of human nature /.../ Perhaps, we can push our enquiries no farther, or pretend to give the cause of this cause" (Hume 1748/2004, 32).

According to Hume, even in the case of the movements of one's own body, we have no direct perception of the cause for those movements (our own power or will): "The powers, by which bodies operate, are entirely unknown. We perceive only their sensible qualities" (Hume 1739/2000, 411). The perception of these qualities is based solely on our past experience. This experience is very close to the passivisation of users that is often attributed to algorithmic personalisation. Events come and go in Hume's world the same way content is recommended on YouTube: any subjective activity is, secondarily, transformed into data, which passivises users and deprives them of their own agency. Or, as Eli Pariser argued in his famous work on filter bubbles: "Algorithmic induction can lead to a kind of information determinism, in which our past clickstreams entirely decide our future. If we don't erase our web histories, in other words, we may be doomed to repeat them" (Pariser 2011, 75).

The power of habit, which in Hume's thought represents the only connection between the past and the future, manifests itself not only as a form of epistemological scepticism, according to which reason is dethroned at the same level as belief, but also as machinic determinism; whereby repetition in the past determines the expectations in the future. The guiding principle of human nature in Hume's thought thus corresponds with the core principle of the "intelligence" of machines (see Sterrett 2002). Even though some say that the term "machine learning" is misleading, because machine "does not have real cognition" and "only humans can learn" (see

Coeckelbergh 2020, 83), one can see that Hume's vision of the basic functioning (and the principle of learning) of the human mind is not presented as a fully conscious process. On the contrary, Hume may be the first modern philosopher who does not tie philosophical certainty and the possession of the truth to the concept of self-consciousness, as was the case with Descartes or Locke (see Balibar 2013). For Hume, subjectivity is never more or less (self)conscious, or more or less close to immediate and definite certainty: "everything what enters the mind is perception, it is impossible to be anything different for the feeling /.../ This is why we can be wrong even when we are most intimately conscious" (Hume 1739/2000, 127). Habit functions automatically, without the subject's conscious awareness and control – this functioning is perceived as a "black box" from the perspective of the human subject. Habits are like machine learning: they don't ask why, they simply aim to automate and economise processes and recognise patterns. They "learn" in the sense that they can serve the purpose of being applied to different similar problems in the future.

We can see how this specific combination of Hume's determinism and scepticism also leads to immoralism. From an empiricist perspective, both moral and immoral acts appear to be on the same level: they are all predictable and conditioned by experience. Hume distances his "experience-based science of human nature" from the natural sciences based on experiment (Norton 2000, I14–I17), but nevertheless he preserves the parallelism between both fields: experience replaces the methodological role of the experiment in predicting human behaviour. Our passions, for example, are essentially related to our experiences of pleasure and pain (Ibid. I46–I50), which implies that those sensation define even higher moral feelings: "pleasure is the 'very essence' of virtue, beauty, and wit" (Ibid. I53; see also Génova and Navarro 2018).

Not surprisingly, Kant rejects this kind of understanding of morality, which is dependent only on previous experiences. In Hume's theory, it is (according to Kant) impossible to locate an instance of a subject's autonomous activity, which is why a seemingly moral action is derived from previous experience (and in the strive for happiness) in exactly the same manner as an immoral one. To answer this problem, Kant's moral action is perceived as pure, and therefore as an action that is completely independent from experience and that is dictated only by pure reason. In his *Critique of the Practical Reason* (1788/2002, 34), Kant rejects "the principle of private happiness" and shows how "the law of the pure will – which is free – places the will in a sphere entirely different from the empirical one".

A moral act that follows the law of pure will could, therefore, be understood as a non-habitual act or as an act that resists habit. The ability to restrain, resist, and control habits and impulses has always been at the core of ethics, but Kant may be the first who systematically confronted moral acts (acts that follow a categorical imperative) and acts performed out of habit. A similar form is taken today by appeals on how individual users can limit the power of algorithms over their lives: They must count on their pure will and the power to resist impulses and stimuli, and on the power to break free from the Humean world of empirical determinism, and to autonomously direct their own attention and their own actions. To paraphrase a typical book on the problems of algorithmically induced biases: One needs a moti-

vation to get rid of the motivated reasoning, i.e. the “motivation to see things as they are (not as we wish them to be)” (Galef 2021, 7).

However, as Bergson (2014, 372–373) argued, Kant’s categorical imperative with its form of a pure obligation (“you must because you must”) itself takes a form that is not far away from the rigidity and repeatability of habit. It is also not completely clear if Kant’s way of rejecting Hume’s “machinic” understanding of human nature is entirely successful, since it establishes the ideal of morality, which remains completely untouched by any empirical instance. Paradoxically, such “cold” morality, even if (or exactly because) dictated by a pure mind, itself seems somehow robotic and rigid. For if all empirical aspects must be excluded from moral decision-making, then also all sympathy and compassion learned through life events, which are often understood as the foundations of human morality, would seem to be lost too.

5. The discrepancy between Kantian ethics and empirical practice

The question we must ask is whether this kind of “frontal attack” on empiricism (from the Kantian position of non-empirical purity) presents a fruitful approach for considering today’s AI ethics, or whether the opposite is true, and – because of its excessive “purity” – such an approach could reproduce the gap between empirical reality and ethical ideals?

As Deleuze remarked, Hume’s empiricism is a sort of science fiction in which one has an impression of a strange, foreign world, “seen by other creatures”, but also “the presentiment that this world is already ours, and those creatures, ourselves” (Deleuze 2001, 35). This uncanny and strange impression may be the main reason why our common sense often joins the Kantian critique of Hume’s empiricism. However, even if Hume’s notion often seem like a strange mixture of scepticism, immoralism, and determinism, which is at odds with the common sense, it works in practice. Empiricism is empirically evident: it draws its strength from its capacity for verification and confirmation. “The result is a great conversion of theory to practice” (Ibid. 36), wrote Deleuze. Even if, according to Hume, it is not possible to predict what will happen tomorrow, it is possible to predict what people, according to the experience through which they have been formed, will believe about what will happen tomorrow. The concept of belief should therefore be understood as “a mental state with a certain dynamic role in the production of action” (Bell 2009, 37). It is exactly the ability to predict such beliefs that is today seen at the very core of the “habitual new media” and the machine learning algorithms (see Chun 2016; Kant 2020), which are, as it is empirically evident, actually quite good at predicting the behaviours, beliefs, and the needs of users.

In contrast, the existence of pure reason or pure will is not so empirically self-evident. This is, unfortunately, structurally a necessary result of the fundamental orientation of Kantian critical thought, which establishes a purified sphere that is separated and independent from anything empirical; whereby the more the critical thought is at odds with empiricism and the less power it has over the course of changing the world, the more it becomes invested in its own criticality. Everything it

criticizes is worth criticizing and – because critique is empirically ineffective – more critique is needed. As Krašovec (2018) wrote (in a slightly different context) the basic structural problem of humanistic ethical theory is that it exists institutionally and habitually separated from the concrete technical organisation of the world, that it is not involved in it, and that its basic premise (and at the same time also the reason for its existence) is precisely this separation.

In this context, it sometimes seems that the demands for AI ethics also represent a complete opposition to the very essence of the functioning of AI, which cannot be understood like any other technological tool that can be used and easily controlled, but must, by its own definition, demonstrate a certain autonomy of non-human action; that is, a capacity to learn and to reason. As Héder (2021, 120) argued, the current wave of artificial intelligence ethics guidelines “can be understood as desperate attempts to achieve social control over a technology that appears to be as autonomous as no other”, indicating a misunderstanding of the very nature of artificial intelligence, which is inherently “resistant to such control”.

Kantian and humanistic predispositions of contemporary AI ethics that emphasise the role of a “human-rights-driven approach” as “the key to building trustworthy AI” (Sorbán 2021) have undoubtedly contributed to the proper addressing of many of the problems posed by this and other such new technologies. However, they are also co-responsible for some of the shortcomings and failures in the regulation of the use of AI. As Vică, Voinea, and Uszkai (2021, 88) argued, “ethics [are] seldom ‘pure’, that is based solely on supreme or ultimate moral principles”. Therefore, Kantian ethical predispositions may not always prove applicable to concrete situations. They can even function as an entrenchment of the status quo. “Cookie consent” notifications may be a good example in this context, since they are – in opposition with the initial intent to control the use of personal data – often “designed to maximise compliance” and treated “as one more persuasive interaction” (Williams 2018, 116).

Kantian foundations could be found, for example, in the approach of the European Union towards data-handling legislation, which it has based on the idea of informed consent to the collection of data. However, such an approach that counts on the user’s ability to reject data collection sometimes does not consider that any kind of “data diet” is almost completely impossible in the era of “data-hungry” algorithms. The Kantian vision of the subject’s independence from all that is empirical and/or from such data-gathering is nowadays difficult to ensure also because it significantly compromises the user experience. Sometimes the principles in ethical handbooks seem so high-flying that they are quite impossible to apply in reality, which is why some argue that the question of “enforceability” – how ethical principles can “become binding” – “should be taken into account more seriously” (Gyulai and Ujlaki 2021, 30). This divergence from empirical reality, which is also creating a gap between what is realistically possible and what is ethical, must be understood as a structurally necessary result of the excessive moral rigour arising from the categorical character of Kantian ethics.

A problem of the Kantian anti-empiricist position is also that it presupposes that there is something immoral and wrong with empirical data in itself, with the process of learning through experience, or with the habitual automation and simplifica-

tion of repetitive actions. On the contrary, I believe that it would be easy to imagine a situation in which such approaches could be used in a manner that could reduce compulsive and addictive behaviour patterns or human biases. Some such methods – for example, Twitter’s warning when a user tries to share content that he or she has not yet read – are already in use. We therefore need to be more specific and empirical in our critique of datafication and algorithmisation. We also need to ask when, how, and why exactly the market logic of the attention economy came to dominate the Internet? Why is data being used mainly for the purposes of attracting users? Why are users – rather than rational citizens who are supposed to participate in public online interactions or as learning subjects who would like to enhance their cognition – being treated as consumers whose needs must be fulfilled?

6. Conclusion: towards the critical empiricism

Let us summarise: The Kantian confrontation with Hume’s empiricism is today manifested as a confrontation between the ethical, normative, and regulatory approaches to technological novelties on the one hand, and the unprecedented rise of big data, algorithmisation, and the use of artificial intelligence on the other hand. According to (Kantian) normative ethics, the use of empirical data supposedly alienates humans from their authentic essence on several levels: our attention is stolen and given over to sensationalist content that appeals to the passions rather than to a reasonable contemplation (in a similar manner as Hume’s theory devaluates reason at the same level as belief); our morality is threatened by empirical stimuli that seek to exploit our momentary desires (in the same way as Kant’s morality is seduced by passions and instincts); our autonomy is replaced by the autonomous and automatic functioning of machinic intelligence (in the same way as there is no space for the concept of autonomous decision in Hume’s empiricism). The “continuous tracking of human life” is, therefore, as many argue, “incompatible with the minimal integrity of the self that underlies autonomy and freedom in all their forms” (Coudry and Mejias 2019, xv).

Although in this article I acknowledge the analytical value of this reasoning, as well as the value of the (Kantian) ethics of artificial intelligence on which the regulation of new technologies is often based, I also note there are certain problematic patterns raised by this kind of frontal opposition to empiricism. Such critique is often too general and too abstract – it is non-productive to oppose data, numbers, and quantity in general or to idealise a romantic image of human nature that transcends material and empirical circumstances. This kind of logic is partly responsible for the gap between normative ethical principles and empirical reality. Even if much of the philosophical understanding of technology in the 20th century was based on such ethical principles – from Heidegger’s rejection of cybernetics to Adorno’s distrust of mass media – today’s technologies based on machine learning and big data seem to require an upgrade of their ethical groundings.

This is why I argue in favour of critical empiricism – a theory that must meet two criteria: 1) it must be able identify the specific empirical actors and their interests

that ultimately reproduce data colonialism, communication inequalities, and digital exploitation in the online sphere; 2) it must be able to interfere in the empirical development of new technologies (and not only oppose this in its empirical totality).

Therefore, Hume's empiricism (similar to today's turn to big data and algorithmisation based on the concept of habit) does not have to be completely dismissed. We should not forget that Hume's sceptic thoughts always left open the principled possibility for different and new ideas to form and actions to occur in the future: perhaps the sun will not rise tomorrow, since it is merely a habit of the mind to conclude that tomorrow will be the same as yesterday. And it was exactly this openness to the future and to novelty that inspired Hume's interpreter Gilles Deleuze, who centred his philosophical *oeuvre* (and his practical ethics) around questions of difference, diversity, and novelty. "It is perhaps habit which manages to 'draw' something new from repetition" (DR, 9), he argued. Habit contracts all that is repetitive and identical, and in doing so it exposes the different, deviant, and heterogeneous. However, habit should not be understood only as that conservative power which reproduces the same, but also as a power that can produce novelty. As Malabou (2008, viii) argued, we should bring to light that habit in the history of philosophy does not represent only "addiction (machining repetition)", but also "grace (ease, facility, power)" – "it is one and the same force, one and the same principle".

Allow me a brief subjective opinion to conclude. Perhaps it is exactly the attempts of the attention economy to control our habits that have given rise to the rebellion against it in the form of a desire for a different, more diverse, and less monotonous internet of the future. Those desires are manifested, I believe, also in the formation of many underground internet movements and local communities of practice, such as those presented by the open-source initiatives. And maybe what is missing is not only the transcendental normative principles of (Kantian) ethics but also more support for practical and immanent attempts as part of the struggle for a free internet and for the democratic use of new technologies. There is no doubt that some of the problems with our habituation to new technologies should be solved in a Kantian manner – with a law, regulation, prohibition, or with withdrawal from the empirical data – however, we should also take into consideration the formative power of habit, which can help with the concrete application of ethical principles in empirical reality and produce something new through repetitive practical activities.

References

- Ariely, Dan. *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins Publishers, 2008.
- Balibar, Etienne (ed.). *Identity and Difference: John Locke and the Invention of Consciousness*. London: Verso, 2013.
- Bell, Jeffrey. *Deleuze's Hume: Philosophy, Culture and the Scottish Enlightenment*. Edinburgh: Edinburgh University Press, 2009.

-
- Bergson, Henri. *Key Writings*. New York in London: Bloomsbury, 2014.
- Carlson, Matt. "Automating Judgment? Algorithmic Judgment, News Knowledge, and Journalistic Professionalism." *New Media & Society* 20, no. 5 (2018): 1755–1772.
<https://doi.org/10.1177/1461444817706684>
- Cera, Agostino. "Beyond the Empirical Turn: Elements for an Ontology of Engineering." *Információs Társadalom* XX, no. 4 (2020): 74–89.
<https://dx.doi.org/10.22503/inftars.XX.2020.4.6>
- Cheney-Lippold, John. *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: NYU Press, 2017.
- Chandler, David. "Digital Governance in the Anthropocene: The Rise of the Correlational Machine." In *Digital Objects, Digital Subjects: Interdisciplinary Perspectives on Capitalism, Labour and Politics in the Age of Big Data*, edited by David Chandler and Cristian Fuchs, 23–42. London: University of Westminster Press, 2019.
<https://doi.org/10.25969/mediarep/11926>
- Chun, Wendy Hui Kyong. *Updating to Remain the Same: Habitual New Media*. Cambridge and London: The MIT Press, 2016.
- Coeckelbergh, Mark. *AI Ethics*. Cambridge, MA: The MIT Press, 2020.
- Coeckelbergh, Mark. "The Tragedy of the Master: Automation, Vulnerability, and Distance." *Ethics and Information Technology* 17, (2015): 219–229.
<https://doi.org/10.1007/s10676-015-9377-6>
- Couldry, Nick, and Ulises A. Mejias. *The Cost of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, California: Stanford University Press, 2019.
- Deleuze, Gilles. *Kant's Critical Philosophy*. London in New York: Continuum, 1967/2008.
- Deleuze, Gilles. *Empiricism and Subjectivity*. New York: Columbia University Press, 1977/1991.
- Deleuze, Gilles. *Pure Immanence: Essays on A Life*. New York: Zone Books, 2001.
- Eyal, Nir. *Hooked: How to Build Habit-Forming Products*. Harmondsworth: Portfolio Books, 2014.
- Galef, Julia. *The Scout Mindset: Why Some People See Things Clearly and Others Don't*. New York: Portfolio/Penguin, 2021.
- Génova, Gonzalo, and Ignacio Quintanilla Navarro. "Are Human Beings Humean Robots?" *Journal of Experimental & Theoretical Artificial Intelligence* 30, no. 1 (2018) 177186.
<https://doi.org/10.1080/0952813X.2017.1409279>
- Gyulai, Attila, and Anna Ujlaki. "The political AI: A realist account of AI regulation." *Információs Társadalom* XXI, no. 2 (2021): 29–42.
<https://dx.doi.org/10.22503/inftars.XXI.2021.2.3>
- Héder, Mihály. "AI and the resurrection of Technological Determinism." *Információs Társadalom* XXI, no. 2 (2021): 119–130.
<https://dx.doi.org/10.22503/inftars.XXI.2021.2.8>
- Hosseinpour, Hesam. "Disobedience of AI: Threat or promise." *Információs Társadalom* XX, no. 4 (2020): 48–56.
<https://dx.doi.org/10.22503/inftars.XX.2020.4.4>
- Hume, David. *A Treatise of Human nature*. Oxford and New York: Oxford University Press, 1739/2000.
- Hume, David. *Enquiry Concerning Human Understanding*. Oxford: Oxford University Press, 1748/2004.

- Kaluža, Jernej. "Habitual Generation of Filter Bubbles: Why is Algorithmic Personalisation Problematic for the Democratic Public Sphere?" *Javnost - The Public* 29, no. 3 (2022a): 267–283.
<https://doi.org/10.1080/13183222.2021.2003052>
- Kaluža, Jernej. "Far-reaching effects of the filter bubble, the most notorious metaphor in media studies." *AI & Society* (2022b).
<https://doi.org/10.1007/s00146-022-01399-x>
- Kant, Immanuel. *Critique of Pure Reason (Unified Edition)*. Cambridge: Hackett Publishing Company, 1781/1996.
- Kant, Immanuel. *Prolegomena to Any Future Metaphysics*. Cambridge: Cambridge University Press, 1783/2004.
- Kant, Immanuel. *Critique of the Practical Reason*. Cambridge: Hackett Publishing Company, 1788/2002.
- Kant, Tanya. *Making it Personal: Algorithmic Personalization, Identity, and Everyday Life*. New York: Oxford University Press, 2020.
- Krašovec Primož. "Spremna Beseda." In *Molekularno rdeče: teorija za dobo antropocena*, by Mackenzie Wark, 311–320. Ljubljana: Sophia, 2018.
- Lewandowsky, Stephen, Laura Smillie, David Garcia, Ralph Hertwig, Jim Weatherall, Stefanie Egidy, Ronald E. Robertson, Cailin O'Connor, Anastasia Kozyreva, Philipp Lorenz-Spreen, Yannic Blaschke, and Mark Leiser. *Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making*. Luxembourg: Publications Office of the European Union, 2020.
- Malabou, Catherine. *The Future of Hegel: Plasticity, Temporality and Dialectic*. London and New York: Routledge, 2004.
- Malabou, Catherine. "Addiction and Grace: Preface to Félix Ravaisson's *Of Habit*." In *Of Habit*, edited by Ravaisson Félix, vii – xx. New York in London: Continuum, 2008.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.
- Myllylahti, Meria. "Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms." *Digital Journalism* 8, no. 5 (2020): 567–575.
<https://doi.org/10.1080/21670811.2019.1691926>
- Norton, David Fate. "Editor's Introduction." In *David Hume: A Treatise of Human nature*, edited by David Fate Norton and Mary Norton. Oxford and New York: Oxford University Press, 2000.
- Totschnig, Wolfhart. "The Problem of Superintelligence: Political, Not Technological." *AI & SOCIETY* 34, no. 4 (2019): 907–920.
<https://doi.org/10.1007/s00146-017-0753-0>
- Tufekci, Zeynep. "Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency." *Journal on Telecommunications and High Technology Law* 13 (2015): 203–217.
- Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Random House, 2011.
- Seaver, Nick. "Captivating algorithms: Recommender systems as traps." *Journal of Material Culture* 24, no. 4 (2019): 421–436.
<https://doi.org/10.1177/1359183518820366>

-
- Slaček Brlek, Aleksader Sašo. "Audience Metrics as a Decision-Making Factor in Slovene Online News Organizations." In *Technologies of Labour and the Politics of Contradiction* (book series *Dynamics of Virtual Work*), edited by Bilić Paško, Primorac Jaka, and Bjarki Valtýsson, 211–232. Cham, Switzerland: Palgrave Macmillan, 2018.
- Sorbán, Kinga. "Ethical and legal implications of using AI-powered recommendation systems in streaming services." *Információs Társadalom XXI*, no. 2 (2021): 63–82.
<https://dx.doi.org/10.22503/inftars.XXI.2021.2.5>
- Splichal, Slavko. *Datafication of Public Opinion and the Public Sphere: How Extraction Replaced Expression of Opinion*. London: Anthem Press, 2022.
- Sterrett, Susan G. "Too many instincts: contrasting philosophical views on intelligence in humans and non-humans." *Journal of Experimental & Theoretical Artificial Intelligence* 14, no. 1 (2002): 39–60.
<https://doi.org/10.1080/09528130210153505>
- Vică, Constantin, Cristina Voinea, and Radu Uszkai. "The emperor is naked: Moral diplomacies and the ethics of AI." *Információs Társadalom XXI*, no. 2 (2021): 83–96.
<https://dx.doi.org/10.22503/inftars.XXI.2021.2.6>
- Vobič, Igor, Marko Robnik Šikonja, and Monika Kalin Golob. "Nazaj v prihodnost: avtomatizacija in preobrazba novinarske epistemologije [Back to the Future: Automation and the Transformation of Journalism Epistemology]." *Javnost - The Public* 26, Sup. 1 (2019): 41–61.
<https://doi.org/10.1080/13183222.2019.1696600>
- Voinea, Cristina, Constantin Vică, Emilian Mihailov, and Julian Savulescu. "The Internet as Cognitive Enhancement." *Science and Engineering Ethics* 26 (2020): 2345–2362.
<https://doi.org/10.1007/s11948-020-00210-8>
- Williams, James. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press, 2018.

Self-protective versus utilitarian autonomous vehicles

The following examination focuses on the moral dilemmas surrounding the comparison of self-protective and utilitarian autonomous vehicles. These vehicles can be programmed to prioritize the safety of passengers in an accident or prioritize the greater good to save more lives. The essay will explore various ethical questions such as evaluating the numbers game approach, analyzing the principles of beneficence correlated with social inequality, and interpreting the principle of autonomy in the context of autonomous vehicles. Additionally, the examination will consider the harm–benefit ratio. In the sense of methods, this analysis uses the classical issues in a new way and provides recommendations for decision-makers to consider.

Keywords: *self-driving cars, utilitarianism, self-protective, moral issues*

Author Information

Beáta Laki, University of Pécs Medical School, Department of Behavioural Sciences
<https://orcid.org/0000-0002-6348-8483>

How to cite this article:

Laki, Beáta. “Self-protective versus utilitarian autonomous vehicles”.
Információs Társadalom XXIII, no. 2 (2023): 83–94.
== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.5> ==

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

In my following essay I aim to explore the ethical dilemmas surrounding the comparison of self-protective and utilitarian autonomous vehicles (AVs) in a unique way. Autonomous cars can be programmed or taught to be self-protective vehicles, when the highest rule to follow is the safety of the passenger in case of an accident situation or through utilitarian approach to serve the highest good and save as many lives as possible.

There are ongoing debates about which of these we should use and choose as owners of AVs. We are faced with many arguments in favor of and against each. But at least one fact is sure: the consequence of introducing autonomous cars into global traffic saves many lives. Because of the nature of AVs, they are able to quickly calculate the best decision in an emergency situation, whereas human drivers may not be able to do so; humans are not capable of thinking through all of the possible scenarios and acting in the best way in such an urgent situation.

This raises questions about how we should choose between the two types of programmable “act” as drivers and as pedestrians? Should we merely play the numbers game or is something else essential? What is the best moral decision, if such a decision exists at all? Is postponing introducing AVs not an immoral behavior by itself?

In this essay I will circle around these kinds of question, taking into consideration, inter alia, the harm–benefit ratio, personal autonomy, the possible wider effect on the whole of society, the lives that could be saved with and without AVs, and issues of social and personal responsibility.

2. Why do we use AVs’?

The primary goal of using AVs is safety and crash prevention, but this has proven to be a complex issue. Before delving into the details, it is worth noting the various benefits of using these machines in everyday life. (Maurer et al. 2016)

Without exhaustive list of benefits, AVs make traveling more comfortable and smooth. They decrease traffic violations. They also provide more safety than traditional driving due to the lack of driver fatigue or impairment: passengers do not have to be afraid of the driver becoming tired, or being under the influence of alcohol or drugs. And if we compare the reaction times, we can find significant difference also: AVs have faster reaction times than humans, making them better equipped to handle unexpected situations (Braun et al. 2020). The environmental effects of AVs are also less due to the system being able to choose the driving mode that uses only as much fuel/energy as necessary; they do not waste fuel on useless fast-speed acceleration, etc.

¹ In this essay when I mention AVs I understand them to be 4th level – high automation – and higher. The 5th-level – full automation – AVs are those that do not need a human to intervene in the driving; they are capable of driving long distances without human support. They are real self-driving cars (Britton 2020).

Despite these benefits, a significant problem remains with the use of AVs: how to protect and make decisions in the event of an accident. How do we have to protect at all? And how to decide to choose one or the other participant of the accident who will be the victim if the harmful event is inevitable?

The book *Moral Machines* (Wallach and Allen 2009) concludes that as autonomous systems become more prevalent in our society, it is important to consider the ethical implications of their actions. The authors argue that current approaches to programming autonomous systems, such as rule-based systems and utility-based systems, are not sufficient for dealing with the complex moral dilemmas that autonomous systems may face. They propose the use of a “hybrid approach” that combines elements of rule-based and utility-based systems with a more comprehensive understanding of ethical principles. In summary, the book concludes that autonomous systems have the potential to greatly benefit society, but that there are also significant ethical challenges that must be addressed in order to ensure that these systems are used in a responsible and ethical manner.

It seems, then, that moral theories may not be that helpful when it comes to real-time decision procedures, since none of them can solve the situation beneficially for everyone. In the following I will analyze this issue from different perspectives and with the help of some arguments of applied ethics.

2.1. An unsolvable problem?

When creating regulations for AVs, two main interests must be taken into account. One is the interest of buyers who want an AV that protects them and saves their lives in case of an accident, while also serving their needs and making their life easier. The other is the interest of pedestrians who want to experience the benefits of AVs in terms of protection and safety. Both of these groups want the same thing: protection and safety. However, it is not always possible to save all lives in the event of an accident, which creates a difficult problem in terms of regulation. At the same time, it is important to create regulations that are acceptable to all parties, regardless of whether they are AV owners or pedestrians.

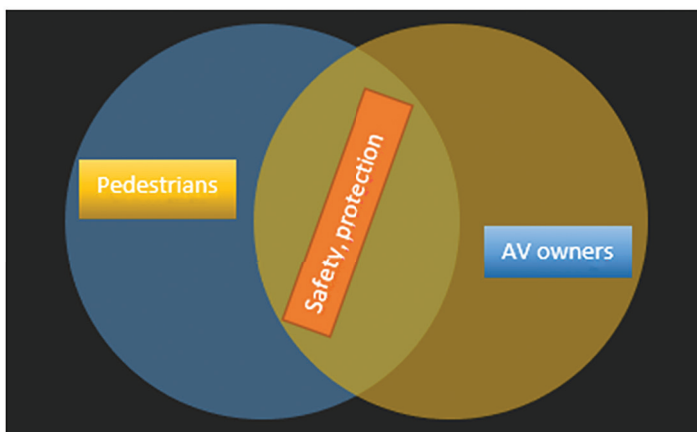


Figure 1. A common cross-section (own creation)

2.2. Self-protective versus utilitarian AVs

We could ask the question why are we talking again about this comparison, self-protective versus utilitarian AVs? The answer is quite simple: because the interests of users and non-users can be argued with these theories.

Self-protective represents a mostly deontological approach that focuses on the passenger's protection. At the base of it is the owner's interest. Otherwise, it is ethically questionable at the same time if we look at the fact that this protective advantage is a privilege of wealthy people since these vehicles cost a fortune. About this we will talk more a bit later in connection with a specific principle.

Utilitarianism approaches the issue from the theory that the highest goal is saving as much life as possible. This depicts the consequentialist aspect and here we are faced with the so-called numbers game. The community interest and the bigger good, less harm are in the center of this and the algorithm operates according to the harm-benefit ratio. The numbers are more important than the use of AVs. At this point it is essential to describe what rule utilitarianism means since this is that type of consequentialist theory that is suitable for this situation. It is a type of utilitarianism that focuses on the rules or principles that govern moral behavior, rather than the outcomes of specific actions. (Miller 2014) According to rule utilitarians, moral rules or principles should be evaluated based on their overall utility or usefulness in promoting the general happiness or well-being of society. The basic idea is that certain moral rules, such as "do not lie" and "do not steal," tend to produce more overall happiness in society than would be produced in a society without such rules. So, it is believed that following these moral rules will lead to the best overall outcome for society as a whole.

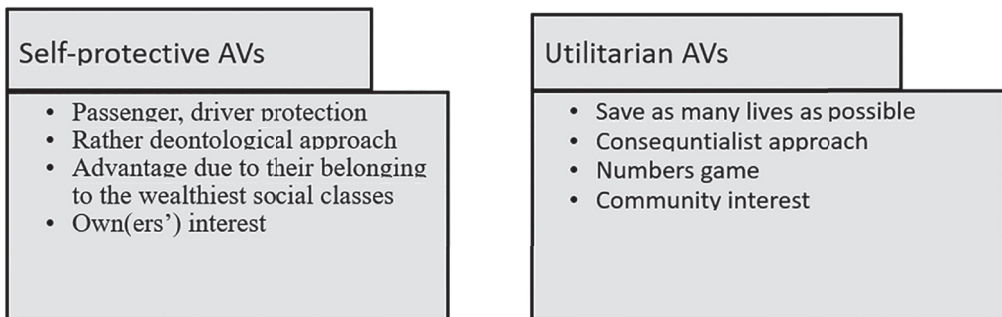


Figure 2. Main differences between Self-protective versus utilitarian AVs

The utilitarian approach can be strengthened by highlighting a few of Patrick Lin's (Lin 2015, 2013a, 2013b, 2013c, 2023) works. He implies that crash optimization, which causes fewer severe injuries, can be applied when programming AVs and this means the as much as it is possible reduction in the severity of consequences. This also follows the consequentialist ethical theory. Crash optimization is driven by a targeting algorithm that calculated the least-victims scenario and with this saves

the most lives. But he also says that crash optimization is not enough since accident situations are complex, different and frequently unpredictable such that we need something else, too, not just this algorithm:

Even if consequentialism is the best ethical theory and the car's moral calculations are correct, the problem may not be with the ethics but with a lack of discussion about ethics. Industry, therefore, may do well to have such a discussion and set expectations with the public. Users – and news headlines – may likely be more forgiving if it is explained in advance that self-sacrifice may be a justified feature, not a bug. (Lin 2015, 77)

Lin also draws attention to the different levels of moral obligation that arise from the ownership of AVs, whether they are publicly or privately owned. Due to it, it would be necessary to formulate distinguishable levels of moral – if we are allowed to say so – obligations of the programmed autonomous car.

3. Trust in it or not? Who or what do we trust?

Until we reach at least the 4th level of AVs – which is in progress with promising results – there are more participants in the decision-making and responsibility sides of the AV issues. Of course, we have to deal with this only if something happens, an accident with personal or property damage. Currently, those accountable can be the owners, the producers, and the software developers/engineers (companies). After we cross the line of human-assisted AVs, the 3rd, “conditional automation” level, only the owners would be accountable because any issues will be down to incorrect maintenance, that is, human error is punished. So, improving AVs decreases the number of accountable parties. Theoretically it makes the whole responsibility process simpler, but practically it does not. Why not?

Because programming AVs contains such tasks as are strongly connected with life–death decision-making procedures. How can we implement into a vehicle a moral way of thinking? Are we allowed to choose one quite acceptable scenario for life-saving situations or do we have to accept teaching self-driving cars to do their best according to computing processes? (Service 2021; Gunkel 2018) And here the unsolvable question “whose interest is in the focus?” appears again.

When the 4th level of AVs is achieved, the human factor loses its importance concerning responsibility; humans remain only in the frame of consumers who enjoy this kind of driving service. In this phase of AVs, since machines, algorithms are not able to make decisions, not even moral ones, on their own because of the nature of their operation. They are artificial tools that are programmed and built by humans. (Z. Karvalics 2015) But accidents can happen, and we have to define the responsible ones. This is an extremely complex and difficult task since the implemented programs and algorithms operate in deep learning systems that we are not able to follow in all cases, although we created the programs that enable them to handle all circumstances.

My current topic does not deal with all the decision-making processes but focuses on the moral decisions. Because of the nature of self-driving car programs, we cannot attribute to them the capability of thinking in such a moral way as humans,

not even by using a high level of artificial intelligence (AI) because it can only be imitation, following patterns and statistics but not real consideration of dilemmas. So, it can be said in this interpretation that self-driving cars are only tools that operate according to our programmed rules. (Pokol 2017) And the trust we put in these machines is misleading. It should be based on trust toward producers and programmers, but what rules and theories do they follow when creating the systems that are suitable for fulfilling the requirements of safe AVs?

4. An ideal scenario and the principles

4.1. *An ideal scenario*

Assume that AVs have no bugs, technical malfunctions, or issues with AI programming and deep learning mechanisms, and are functioning perfectly (although this is nearly impossible; we will just imagine that it is so, for the purpose of this thought experiment). In this scenario, we can eliminate the risk of car-related issues. However, accidents may still occur due to human factors, specifically the creators of the AVs. At this point, the principles of consequentialism and deontology come into play and may conflict with each other. In the following analysis, using this ideal mechanical background, I will examine principles that can both strengthen and weaken the introduction of AVs. Although, my goal is not to leave issues unresolved, it is important to recognize that there may be irresolvable contradictions and that an acceptable solution, rather than a perfect one, must be considered when weighing the opportunities and drawbacks.

The principle of justice may be compromised to some extent, but this compromise may not be fully acceptable. This can be compared to the intentions of the principle of beneficence. I will argue that the principle of beneficence is stronger than the principle of justice, but further examination is necessary to understand how this is the case.

4.2. *Principle of justice and social inequality*

There is no need for special knowledge about the problems of justice and social inequality in the world since we meet them in our everyday lives. This topic is another element on the list that does not decrease current societal issues but, rather, increases them. As I mentioned earlier, since AVs cost a fortune, it is not possible for everybody who has a car to have one. This strengthens inequality and makes the issue of “benevolence” more complex.

The problem with the price, if these can only be bought/owned by wealthier people, is that it won't be morally manageable to make their safety higher in contrast with the safety of poorer people. This privilege is neither acceptable nor fair.

It's true that more protective, higher-quality tools are more expensive, but in the case of AVs this protection isn't only passive; it is an active intervention that can

also cause harm – not intentionally but as its consequence – when AVs try to avoid dangerous situations. The question arises, can one buy the security of one's life? The answer is yes, if AVs are available, but from a moral perspective, it is not fair if only the wealthy can afford them. To ensure justice and decrease the previously mentioned factor, AVs should be made accessible to everyone. Furthermore, it would be morally necessary to act like this if the main intention of the society was to decrease accidents, especially lethal ones. It would be an obligation and not a choice. At the same time, it means a kind of force of desirable owners/users to have AVs instead of general cars if the safety and the amount of saved lives are in the center of the regulation.

According to the doctrine of double effect, we can accept the loss because our intention and the consequence differ from each other. But if loss means that we can be victims, the situation cannot be attractive enough. A real moral dilemma appears here, also, that is inevitable: lethal car accidents happen and we cannot cease them completely. It is not possible to act perfectly right because one or other influential factor cannot be changed just handled somehow. Due to the occurrence of safely unsolvable situation, somebody is going to be harmed. We have the ability to decrease the risks of accidents, but we have to define who will be the victims in those cases when somebody has to die. At this point the principle of justice could be involved, but actually it is not capable of helping in those decision-making processes. We could ask: are we allowed to decide who has the right to live and who should die? Not really, but the point is we have to formulate scenarios and rules, taking into consideration the consequences. And if we deal with the inequality factor in the society and the advantage derived from this, we are still in a moral trap.

The message of self-protective AVs seems to be that the lives of the rich are worth more. It is unquestionable and obvious; it is an obvious negative discrimination toward others. The principle of justice is being harmed by the fact of social inequality.

4.3. Principle of beneficence

Beneficence is a tricky expression in the case of real moral dilemmas since it is a principle that cannot be represented perfectly in certain situations. If we introduce utilitarian AVs and oblige the likely wealthy customers to buy this kind of vehicle, it increases their possible harm and does not accurately represent the basic idea of AVs: to save lives. Because of this fact, if the utilitarian result of a probably lethal accident is the death of the AV owner, it obviously cannot be in the interest of the owners.

Although the frequency of deadly accidents is statistically extremely low, sometimes they occur because of the nature of traffic, the behavior of other people on the move, etc. In all deadly cases it is necessary to make decisions and compromises; it cannot be avoided. We have to take into consideration the interests of all the participants. And not just morally but in general, we are not allowed to favor financial interests over human lives.

The point of the principle of beneficence is, hopefully, that it is possible to make a decision that has the best outcome, and it can be a higher number of saved lives.

This shows the theory behind it: saving as many lives as possible. This so-called numbers game has mathematically the most beneficial result, but otherwise we can take into consideration different factors, e.g. quality of life, age, life expectancy, race, and so on. But this does not lead us to equal, fair, and acceptable decision-making processes.

On the one hand, forcing buyers into taking more risk if they want to have an AV is not morally acceptable. But concerning the previously mentioned numbers game, saving as many lives as possible would have to be our duty because of the higher good that can be achieved through that. On the other hand, if we program AVs as self-protective vehicles it is not fair either. Measuring it morally, it is more problematic since we give the privilege to people because they are wealthier, and they can afford to buy these tools.

The consequence of this approach is that utilitarian programmed self-driving cars can discourage buyers from owning one since – even if the chance of serious accidents is extremely low – they could be harmed if they trust in the machine that should serve their comfort, the user’s interests and needs, as well as the higher good, they can become victims.

In terms of discrimination, too: if you are rich and want to buy such a vehicle, you have to take the risk that you will be sacrificed for the greater good. It is like a compulsion, or better said, a deal: accepting the risk even if it is extremely low, and enjoying the benefits of AVs hoping nothing serious will happen ever.

It can be seen that the principle of justice and beneficence actually harms from both perspectives but on a different level of severity (in relation to buyers, pedestrians, and others who are participants of traffic).

5. Conclusion – and something else

It is unquestionable that everybody has the right to live, but if a real moral dilemma appears we are obliged to decide how can we influence the consequences and what can be our duties. There is no difference between doing something or not – in the interpretation of non-action is an action also – in connection with making decisions, since both of them have certain results. But in the case of AVs, the consequences can be dramatic. This is why it is necessary to find the best solutions for specific events. *“[S]ome crashes will require AVs to make difficult ethical decisions in cases that involve unavoidable harm”* (Bonnefon 2016, 1573).

It is inevitable that we need to think about moral duty: introducing the available technology is a must if it enables making traffic safer. Not using the technology is equal to omission and unnecessary risk. *“Second—and a more serious problem—our results suggest that such regulation could substantially delay the adoption of AVs, which means that the lives saved by making AVs utilitarian may be outnumbered by the deaths caused by delaying the adoption of AVs altogether”* (Bonnefon 2016, 1575–1576).

Through my writing, I highlighted the difficulties of decision-making and the main obstacles, but I did not recommend specific solutions. Although it would be

useful to formulate rules and regulations to handle the morally problematic issues here, I did not do that. That has to be the result of cooperation among certain professional parties, followed by the principles of the main aim: saving lives, making traveling, driving, or just walking safer for everybody. Because of the nature of this kind of decision-making, a less bad solution could turn out to be one of the best ones, since it is not an option to have a perfectly good one.

Accountability and a possible solution?

If we dive into deeper layers and focus on taking risk, it can be said that since introducing AVs is in everybody's interest, it is actually a must. But in this case personal/individual accountability is not an acceptable approach. The consequence of the events is out of our control. Following this way of thinking, it is nonsense to impeach users who do not have any impact on traffic situations. (Bartneck et al. 2021)

Let's think through a logical but theoretical scenario when the interests, needs, and risk-taking factors have been taken into consideration, involving governments as a kind of responsible organization in the following manner. (Tilesch and Hatamleh 2020)

First, it is practical to collect features, needs, and possible consequences of this theoretical case. The collective interest of introducing AVs is the core of the whole situation.² The reason is not new, because of the life-saving benefits of using more-protective tools in everyday life. Maximizing the effectiveness of this traffic opportunity – in the sense of the numbers game – governments should support development processes and make AVs available not just to wealthy people. At this point it is necessary to highlight accountability, which is one of the most essential questions. If we want to introduce AVs because of safety reasons and we do not want to punish users and non-users with the unlucky, rare, and undesired, unwanted consequence of an unfortunate accident, it is more realistic to transfer the financial and other consequences to the organization that “forced,” advertised, and supported the use of these tools.³ The intention of representing AVs is clear: the safety risks are small compared to general cars and drivers and that can drive a theoretically acceptable risk-taking level.⁴

Concerning other effects on the government side, it requires an increasing amount of financial resource investment if equal availability is the focus (Héder 2020). At this point we can formulate another question: should everybody get such a car (AV) who has a general one or who wants to have a car in general? Being prepared for this financially and measuring legitimacy demands are complex issues.

² And it was through the whole essay also, so please forgive me for the repetition, but it is not acceptable to not mention it again right here.

³ Although it is in the society's interest.

⁴ If it will ever have such a level. Since we are talking about human lives, it is not a perfect formulation of the thoughts, but in a comparative sense the fewer victims, the more acceptable and supportable it is. Without comparison, no victim is acceptable ever. But this statement cannot be represented in practice with the current conditions.

Instead of this economic calculation I would recommend a golden mean that could act as a kind of solution for utilitarian versus self-protective self-driving cars: the Autonomous Vehicle-Car Sharing System (in the following: AV-CSS).

In this idea, governments have the responsibility to operate and maintenance AV car fleet. This makes it clear who will be responsible and accountable in accident situations. If somebody wants to use an AV, they have the chance, since it should be affordable, but the user takes the risk that they will become an innocent victim in a lethal accident should the AV try to act according to consequentialist rules.⁵ On account of the nature of “AV safety” probability, the risk taken is significantly lower than in general cases. The more AVs are participants of public transportation and traffic, the less probability of lethal accidents there is. Because of the kind of common network in relation to knowledge base that teaches AVs about each traffic situation, it makes this kind of transport safer, and it can become more calculated with less risk, so there is less likelihood of harm and in the ideal case, finally, it should be possible to eliminate lethal accidents entirely.

With this AV-CSS, forcing anybody to use an AV is avoidable, but it makes it possible for everybody to live with the opportunity and enjoy the comfort of these tools at the same time. Individual free decision-making opportunity is preserved and represented.

Moreover, car sharing has other beneficial consequences, especially if AVs are electrical cars. Less environmental load, fewer emissions, less expense, fewer cars owned. All of these mentioned results of AV-CSS have a huge impact on society and on the environment as a whole. So, from my point of view, making steps toward this kind of less consumer-centric behavior is worth considering. It is in the interest of all of us. Not by the way, it can solve the analyzed issue of whether self-protective or utilitarian AVs should be introduced? This idea gives a potential way for handling the issue and giving the opportunity to decide freely: to use or not to use. The possibility of choosing the lesser evil should be open to individual decision. What we need is transparency, human oversight, and public and governmental support. Taking into account all of these factors will not solve all of the issues, but with help they can be decreased and narrowed down.

⁵ As has become obvious, this scenario operates according to consequentialist theory – saving as many lives as possible. The reason is that any of the approaches possibly have bad consequences, but in this case the user decides whether they take the relatively low risk and use AVs or not.

References

- Bartneck, Christoph, Christoph Lütge, Alan Wagner, and Sean Welsh. *An Introduction to Ethics in Robotics and AI*. Cham: Springer, 2019.
https://library.oapen.org/viewer/web/viewer.html?file=/bitstream/handle/20.500.12657/41303/2021_Book_AnIntroductionToEthicsInRoboti.pdf?sequence=1&isAllowed=y
- Bonnefon, Jean-Francois, Azim Shariff, and Iyad Rahwan. “The social dilemma of autonomous vehicles.” *Science* 352, no. 6293 (June 24, 2016): 1573–1576.
<https://www.DOI:10.1126/science.aaf2654>
- Braun, Joachim von, Margaret S. Archer, Gregory M. Reichberg, and Marcelo Sánchez Sorondo. *AI and Humanity*. Cham: Springer, 2020.
<https://doi.org/10.1007/978-3-030-54173-6>
- Britton, Jill. “What are the autonomy levels for autonomous vehicles?” Accessed July 4, 2022.
https://www.perforce.com/blog/qac/6-levels-of-autonomous-driving?utm_source=googleadwords&utm_medium=cpc&utm_campaign=QAC-Dynamic&utm_adgroup=Dynamic&gclid=Cj0KCQjwn4qWBhCvARIsAFNAMiudzbtqL699pW39khOyco-3i3mBYr-4v3lO2G_OfOdhgP0e9Kez6UaAg19EALw_wcB
- Héder, Mihály. “A criticism of AI ethics guidelines.” *Információs Társadalom* XX, no. 4 (2020): 57–73.
<https://dx.doi.org/10.22503/inftars.XX.2020.4.5>
- Gunkel, David J. *Robot Rights*. Cambridge, MA: MIT Press, 2018.
- Lin, Patrick. “Why ethics matters for autonomous cars.” In *Autonomes Fahren*, edited by Markus Maurer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner, 69–85. Heidelberg: Springer, 2015.
- Lin, Patrick. “The ethics of saving lives with autonomous cars is far murkier than you think.” *Wired* (July 30, 2013a). Accessed Jan. 4, 2023.
<http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/>
- Lin, Patrick. “The ethics of autonomous cars.” *The Atlantic* (October 8, 2013b). Accessed Jan. 4, 2023.
<http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>
- Lin, Patrick. “Why the drone wars matter for automated cars.” Accessed Jan. 4, 2023.
<http://stanford.io/1je1Quw>
- Maurer, Markus J., Christian Gerdes, Barbara Lenz, and Hermann Winner. *Autonomous Driving*. Heidelberg: Springer Nature, 2016.
- Miller, Dale E. “Rule utilitarianism.” In *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston, and Dale E. Miller, 146–165. Cambridge: Cambridge University Press, 2014.
- Pokol, Béla. “A mesterséges intelligencia.” *Információs Társadalom* XVII, no. 4 (2017): 39–53.
<https://dx.doi.org/10.22503/inftars.XVII.2017.4.3>
- Service, Robert F. “Learning curve, new brain-inspired chips could provide the smarts for autonomous robots and self-driving cars.” *Science* 374, no. 6563 (October 1, 2021): 24–25.
<https://www.science.org/content/article/new-brain-inspired-chips-could-soon-help-power-autonomous-robots-and-self-driving-cars>

-
- Tilesch, George, and Omar Hatamleh. *Between Brains: Taking Back Our Future in the AI Age*. PublishDrive, United States, 2020.
- Wallach, Wendell, and Colin Allen. *Moral Machines: Teaching Robots Right and Wrong*. Oxford: Oxford University Press, 2009.
- Z. Karvalics, László. “Mesterséges intelligencia – a diskurzusok újratervzésének kora.” *Információs Társadalom* XV, no. 4 (2015): 7–41.
<https://dx.doi.org/10.22503/inftars.XV.2015.4.1>

Technics as Hermeneutics

In this paper, I outline, how technics can be understood as a special case of hermeneutics. My argument hinges on the concept of practice, which, according to Joseph Rouse, is hermeneutic activity. I show how Rouse's practical hermeneutics and Don Ihde's material hermeneutics can be considered to be founded on the pragmatist theory of meaning. For John Dewey, interpretation is thoroughly technical.

Keywords: *Philosophy of technology, hermeneutics, Joseph Rouse, Don Ihde, pragmatist theory of meaning*

Acknowledgements

This study was supported by the Estonian Research Council (PRG 462), the Estonian Ministry of Education and Research (IUT 20-5), and by the European Union European Regional Development Fund (Centre of Excellence in Estonian Studies).

Author Information

Juho Lindholm, University of Tartu

<https://orcid.org/0000-0001-6829-295X>

How to cite this article:

Lindholm, Juho. "Technics as Hermeneutics".

Információs Társadalom XXIII, no. 2 (2023): 95–116.

== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.6> ==

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

Hermeneutics first emerged as a discipline for the interpretation of sacred texts in several religious traditions. However, in the 19th century, Friedrich Schleiermacher (1768–1834) enlarged its jurisdiction to all of language. Later, Wilhelm Dilthey (1833–1911) enlarged hermeneutics even more to all meaningful human activity. One of Dilthey’s lifelong projects was to defend the methodological independence of human and social sciences from the imperialist claims of the philosophers of natural science of his time. Roughly, his thesis was that natural sciences operate upon unambiguous observations and uninterpreted formal languages and thus have a claim to objectivity, whereas human and social sciences are necessarily interpretive and depend on the person who makes the interpretation.

For many decades, it was considered that natural sciences can be distinguished from human and social sciences on this basis. Philosophers of natural sciences – the most influential of whom subscribed to logical positivism and empiricism – and philosophers of human and social sciences alike seemed to be happy with such a sharp division of labour. However, the collapse of logical positivism and empiricism, which has taken place gradually since the 1950s, has revealed the existence of profound hermeneutic features in natural sciences.

Ludwig Wittgenstein’s posthumously published later works (e.g. [1953] 2009, 1958, 1969) began to question many tenets of logical positivism and empiricism. In particular, his concepts of *language-game*, *hinge propositions*, and *form of life* are fundamentally at odds with the then-received view in analytic philosophy. His thesis was, in short, that concepts make sense only within a language-game, which consists of language and the actions into which it is woven ([1953] 2009, 8); that knowledge is founded upon hinge propositions; and that language-games and hinge propositions presuppose a form of life. Hence the interpretation of concepts or knowledge requires the adoption of a particular form of life, which qualifies as hermeneutic activity.

According to the Duhem–Quine thesis, observation is theory-laden, or, equivalently, theories are empirically underdetermined (Duhem [1906] 1954; Quine 1951). In the radical interpretation of that thesis, it is not possible to make an observation without presupposing some theory. Hence all observation involves interpretation in terms of a presupposed theory, which qualifies as hermeneutic activity.¹

Michael Polanyi ([1958] 1962, [1966] 1983) questioned the then-prevailing notion that all knowledge could be expressed explicitly and representationally independently of practice, and introduced the notion of *tacit knowledge*. The underlying idea of this is that in order for representational knowledge to be effective, it must be personally translated into practice, which qualifies as hermeneutic activity.

¹ It seems that the theses of Pierre Duhem and W. V. O. Quine are not strictly the same; indeed, their differences may be as great as their similarities. What is common to both though is that they question the independent or self-sufficient determinacy of observations. Thus both lean towards a version of Kantianism: recall that Kant ([1781/7] 1956) argued that observation necessarily involves conceptual activity and hence cannot be considered absolute or objective – as a revelation of *Dinge an sich* (things-in-themselves) independently of how we conceive them.

Thomas Kuhn ([1962] 1996) argued that science evolves in cycles of periods of normal science, the accumulation of anomalies, a crisis, and revolutionary science. In a period of normal science, scientists agree about a paradigm. In a period of revolutionary science, a new paradigm is created, thereby beginning a new period of normal science. Concepts are incommensurable between different paradigms; and the choice between paradigms is arational. Hence Kuhn's concept of *paradigm* comes close to Wittgenstein's *form of life* in the respect that the interpretation of a concept requires the adoption of a particular paradigm.

For decades, the theses of Wittgenstein, Quine, Polanyi, and Kuhn were discussed under different titles, but not as hermeneutics. Joseph Rouse (1987, 26–69) was among the first science and technology studies (STS) scholars to take the hermeneutic character of natural science seriously. He began with a novel interpretation of Kuhn. Traditionally, Kuhn has been held up as a proponent of a theory-driven picture of science. Rouse (1987, 26–40), however, argues that this is a misunderstanding: Kuhn seemed to suggest that science is practice instead, and that “paradigm” denotes shared practice, not shared belief. Rouse contrasted these two ways of understanding Kuhn. Both involve a universal hermeneutics of natural science. The traditional picture of Kuhn is associated with theoretical hermeneutics, while Rouse's alternative interpretation is associated with practical hermeneutics. In theoretical hermeneutics, interpretation takes place strictly within language and consists in the formulation of hypotheses. Practices are evidence for such interpretive hypotheses. On the other hand, practical hermeneutics understands practices themselves as meaningful and significant in their own right and interpretation as one practice among others. In this sense, language need not be involved. Rouse seems to reject theoretical hermeneutics as overintellectual and detached from actual life and its concrete particularities.

Roughly a decade later, Don Ihde (1998, 139–195) took up Rouse's practical hermeneutics and incorporated it within his philosophy of human–technology relations, which could be called *instrumental* or *material phenomenology*. In his earlier works (Ihde 1974; [1977] 2012; 1979; 1990), he had developed Patrick Heelan's idea that using an instrument transforms an experience (Ihde [1977] 2012, 100). Such a transformation shows and magnifies certain properties of objects unobservable to the naked eye. The more experience is transformed, the more it requires interpretation to make sense of it. He also employed Bruno Latour's (1987, 67–68) definition of an *instrument* as an *inscription-producing device*; in Ihde's case, these “inscriptions” are the visual outputs of scientific instruments. Ihde argued that in order to understand the meaning and significance of scientific imaging, interpretation is needed. Because instruments transform experiences, such imaging does not present objects as they appear to the naked eye; if it did, it would be pointless. Hence, without proper interpretation, they would remain unintelligible. However, because of the non-linguistic nature of pictures, the notions of meaning and interpretation must be expanded beyond language.

Both Rouse and Ihde provide important and interesting insights about the interpretation of practices and the instruments they involve, but they do not explain *what exactly* is being interpreted. In other words, they do not explain the *meaning of meaning* they implicitly invoke. Thus there arises a question about what the meanings *are* that underlie practices and instruments.

In this article, I try to fill in that loophole in the defence of Rouse and Ihde. I suggest that the pragmatist theory of meaning, introduced by Charles S. Peirce (1839–1914) and developed further by John Dewey (1859–1952), provides the theoretical basis of practical hermeneutics. In particular, Peirce’s original version of the theory took *non-linguistic meanings*, which practices and instruments seem to exhibit, into account. Peirce even seemed to have modelled scientific inquiry as a special case of *semiosis*, or the process of the interpretation of signs, already by the turn of the 20th century, if not earlier.² Hence, arguably, he anticipated the interpretive or hermeneutic turn in the philosophy of science instituted by Wittgenstein, Quine, Polanyi, and Kuhn, by at least half a century. Dewey, who applied the evolutionary theory to many areas of philosophy, rejected essentialism and replaced it with a study of the conditions of the genesis and development of phenomena – the missing link between evolution and technics (Dewey 1910, 1–19). Thereby he also accounted for the emergence of meaning in the discovery of cause—effect relations through experiment. In short, the pragmatist theory of meaning rejects the reifying notion that meanings are *objects* – least of all *abstract* objects. Rather, they are *potentially enacted*: they are *habits*. But habits are public, observable and causal processes and hence qualify mainly as *objective*, though they may have a subjective facet too (Dewey 1916a, 54–58; 1922, 14–88).

I ultimately conclude that Dewey’s account on the genesis of meaning is *technical* through and through. This suggests that *technics is inherently hermeneutical*.

I begin this paper by briefly citing Rouse’s practical hermeneutics and Ihde’s visualism. I then continue by reviewing the pragmatist theories of meaning and discuss the genesis of meaning together with explaining how they can accommodate Rouse’s and Ihde’s insights. Then, I make some critical remarks, before finally concluding with some suggestions for further inquiry. This article expands on Lindholm (2022).

2. Theoretical and Practical Hermeneutics

Rouse (1987, 41–68) reviewed two ways of accounting for the hermeneutic features of science: theoretical and practical. Arguably, both kinds of universal hermeneutics had already appeared in the works of Dilthey (1956; 1957).

² Peirce seems to have entertained the idea that the acquisition of knowledge in general, and science in particular, takes place in signs already in 1868–9 (see “Questions Concerning Certain Faculties Claimed for Man” (Peirce CP 5.213–263; EP 1, 11–27) of 1868 and its follow-ups “Some Consequences of Four Incapacities” (Peirce CP 5.264–317; EP 1, 28–55) of 1868 and “Grounds of Validity of the Laws of Logic: Further Consequences of Four Incapacities” (Peirce CP 5.318–357; EP 1, 56–82) of 1869). But it is in his mature philosophy and semiotics where that notion comes to fruition. That was crystallized in his 1906–9 distinction between the *dynamical object* and the *immediate object* (which he had called the *primary object* and the *secondary object*, respectively, in 1903; see his original definition in (Peirce CP 2.310–311; EP 2, 275) and its development in (Peirce CP 4.536; EP 2, 407), (Peirce EP 2, 480–481). The latter denotes an object as represented by a sign: how that object appears at any stage of inquiry. The former denotes the real object independently of how the sign represents it: how it appears at the conclusion of the inquiry, defined as the point where everything about the object is known, and no further increase in knowledge is possible. Peirce professed fallibilism from early on (see his 1868–9 papers mentioned above) and hence should have concluded that we cannot possibly determine when that point has been reached. He also should have concluded that the dynamical object might change, which makes it imperative to constantly keep checking whether the immediate object is adequate.

Theoretical hermeneutics takes its departure mainly from Quine's philosophy of language. In theoretical hermeneutics, interpretation takes place within language, it consists of forming hypotheses, and it is analogous with translating a sentence. It denies language-independent access to the world. It takes practices simply as evidence for theories or beliefs. It considers observation statements as hypotheses, and maintains that all hypotheses acquire their meaning in relation to all other hypotheses, or a "web of belief." Hence the difference between observation and theory is merely relative. The aim of theoretical hermeneutics is truth, or accurate representation.³ Theory is understood as disengaged from and independent of practice. That makes experiment and discovery incidental. Only rational reconstruction matters. Observation is merely theory by other means, and not even independent means. This seems to create an unbridgeable gap between language and the world. Rouse suggested that theoretical hermeneutics can be modelled by Mary Hesse's (1980, 125–128) self-programming learning machine. (Rouse 1987, xii, 47–53, 69–72, 98.)

Practical hermeneutics, on the other hand, sets out from Heidegger's notions of being-there (*Dasein*) and being-in-the-world (*In-der-Welt-Sein*) as explained in his *Sein und Zeit* ([1927] 1977). It is analogous with engaging in a practice. It interprets practices themselves. It is not confined to describing and predicting behaviour but also understands the point and significance of doing certain things and not others. Everyday practices interpret the world by adjusting behaviour to the surroundings, by practically responding to them. This need not be representational. Such interpretation takes place in accounting for things: using, avoiding, taking note of, caring for, and discarding them. Learning a skill is learning a field of possibilities. This need not be explicit. Interpretation takes place by presupposing a background, but it is not constituted by hypotheses or theories but rather by ways to be in the world. The background is a configuration of equipment, persons, and the physical setting which is already disposed towards certain possibilities. The understanding thus acquired is not a conceptualization but a performative grasp of how to cope with situations that cannot be abstracted from the world. (Rouse 1987, xii, 47–50, 58–65.)

In theoretical hermeneutics, the interpreter is not committed to how things are. Hence the choice of hypotheses and theories is in principle arbitrary. We can remain indifferent about them. That allows theoretical hermeneutics to abstract from particular conditions. In practical hermeneutics, the situation is quite otherwise. The presuppositions of interpretation are conditions that are not open to deliberation, negotiation, or choice. We cannot renounce them without losing grip. Hence they *matter* to us. (Rouse 1987, 64–65.) Thus Rouse seems to reject theoretical hermeneutics as overintellectual and remote or even detached from actual practical engagement with the world. I would call theoretical hermeneutics an updated version of Cartesianism where the soul is simply replaced with language.

Perhaps the difference between theoretical and practical hermeneutics can be expressed as the difference between determining *what the case is* and what it is to *be*. The latter is rather shown than said. (Rouse 1987, 67–68)

³ Rouse considers Richard Rorty to be a theoretical hermeneuticist, but I believe that Rorty would have opposed the representationalism that Rouse assigns to theoretical hermeneutics.

For the purposes of this article, I define *practical hermeneutics* as the thesis that practices are meaningful in their own right; that a practice consists in a configuration of people, their skills, their activities, their purposes, equipment, and environment; that interpretation is itself a practice; and that the world is interpreted by practically responding to it.

The main upshot of practical hermeneutics is that scientific practices, including experiments, already involve interpretation. Another implication is that the concept of *nature* is ineliminably political: because nature is accessed through practice, what it is to be natural is at stake in practices; what is considered “natural” or “unnatural” has normative significance and is sanctioned by society; and, because nature is experienced as a field of opportunities for action, much of our self-understanding arises from our understanding of nature. The counterfactuals that support causal claims depend on agency; and, conversely, our sense of agency depends on discovered causal relations. Hence nature is not neutral but plastic. Rouse concluded that a practical configuration of the world that involves agents, their practices, their material setting, and what makes sense to do is a necessary condition⁴ for anything to be intelligible at all. (Rouse 1987, 181–185.)

3. Visualism

Don Ihde (1998, 139) argued that science can do a hermeneutics of things by turning them into scientific objects. He continued Rouse’s practical hermeneutics that he considered ontologically oriented, in distinction from the linguistically oriented theoretical hermeneutics (Ihde 1998, 147–148). He believed that the field was clear for introducing his ideas, the purpose of which was “to reconverge what began to diverge with early modernity,” because of recent developments in philosophy of science: logical positivism had dwindled, the sociology of scientific knowledge (SSK) had arisen, and the feminist philosophy of science had emerged (Ihde 1998, 139–150). Ihde is primarily interested in what he calls *visualism* in science – how scientific objects are accessed via the technological construction of images. He argues that the hermeneutics of science is of a special kind: not necessarily linguistic or even propositional but first and foremost *bodily* and *perceptual* – *visual* in particular. Here he follows Husserl and Merleau-Ponty. The scope of scientific imaging ranges from isomorphic to non-isomorphic depictions of scientific objects. The less it retains isomorphism with the object, the more hermeneutic activity is required to make sense of it. (Ihde 1998, 151, 184, 187, 196) The interpretation of visualizations of data is a matter of *learning to see* – to figure out which patterns indicate something else and which do not. Such learning takes place dialectically with the development of the instruments of observation, recording, and visualization. (Ihde 1998, 177–180.)

He outlined a “weak program” that identifies hermeneutic dimensions implicit within current science praxis (Ihde 1998, 151–169) and a “strong program” that ex-

⁴ Rouse did not specify, whether it is a necessary, sufficient, or necessary and sufficient condition. I thank my friend Miika Oksman for pointing this out.

amines the cutting edges of science's knowledge constitution in a hermeneutic way (Ihde 1998, 170–183). He defined the former as “an attempt to reconstruct accounts of science praxis, showing the implicit *hermeneutic practices* already at play within science” (Ihde 1998, 152) and the latter as “potentially more normative” and as “an attempt to push, positively, certain P–H [phenomenological and hermeneutic] practices by way of suggestion and adaptation toward science practice” (Ihde 1998, 152). Ihde's description of the “strong program” and how it differs from the “weak program” were vague. I gather that the “strong program” builds upon the “weak program,” but also extrapolates already existing possibilities into the future.

The “weak program” is partly based on the insights of the earlier, non-hermeneutic tradition itself. Many philosophers of science have divorced science from the lifeworld (*Lebenswelt*),⁵ but Ihde argues that such dualism is false: scientific practices and instruments keep science strictly within the lifeworld. Laboratories and their instruments give a “voice” to things, which yields information about them, by recording measurements in visual outputs like graphs or diagrams. Thereby they make non-visual phenomena visual. But it is not the object measured alone that “speaks”; it is the system of the object and the measurement apparatus. Hence “instrumental artifacts” and calibration errors are always possible and must be accounted for (Ihde 1998, 185–186). In order to determine the part of the object, the part of the instrument can be eliminated by making “instrumental phenomenological variations”: for example, by using different instruments, a multivariant set of measurements, intersubjective checking, deliberate application of focus shifts, and figure/ground reversals (Ihde 1998, 185–186). Due to the effect of instrumentation, the objects are not simply “given” but, rather, made “readable.” That creates the need for the instrument to be as transparent as possible. In an important sense, the resulting images are constructs, but the very effect of a successful construction is emphatically to present the object as authentically as possible. In other words, construction of the image is needed in order to cancel the effects of the environment of the object and of the instrument used. The “weak program” is manifested in imaging technologies such as oscillography, spectrography, intestinal probes, X-ray imaging, ultrasound, sonograms, MRI, fMRI and PET scans, radioactive tracers and dyeing, uranium series dating, carbon 14, electron spin resonance techniques, thermo-luminescence techniques, DNA fingerprinting, MDNA lineage tests, computer-assisted tomography, and radio astronomy. Ihde recognized three trajectories in the “weak program”: optical imaging; non-optical imaging, which begins with X-ray imaging and continues in, for instance, fMRI and PET scans; and microscopy, which continues, for instance, in electron microscopy and radio crystallography. These trajectories retain a degree of isomorphism with the object. The situation changes, however, with oscillographs, spectrographs, charts, graphs, and diagrams. Even mathematics can be made perceptual by depicting curves and surfaces, while computer graphics makes it possible to depict even fractals and chaotic and random phenomena. (Ihde 1998, 151–169)

⁵ For Husserl's concept of *Lebenswelt*, see his ([1936] 1976). He added this notion to his phenomenological philosophy after Heidegger's *Sein und Zeit* ([1927] 1977) without even mentioning his pupil.

As concerns the “strong program,” Ihde extrapolated a number of technological possibilities already present in contemporary instrumentation. As indicated above, the development of visualizations is accompanied by learning to see. One important contemporary technique to assist in such learning is image manipulation – especially zooming in and out, enhancement, contrast, and false colour –, often performed by computers. Some image manipulations even make it possible to convert non-isomorphic images into visual ones. The effect of the atmosphere can be computationally removed from telescope images. Thermal imaging extends human vision into serpentine vision, light amplification into feline vision, and wide-angle imaging into insect vision. Several sources of visual information can be fused into composite images. Computers allow even more imaginary transformations of the visual: avatars, online meetings, computer graphics, virtual and augmented reality, and holography. (Ihde 1998, 170–183, 191–192; [1977] 2012, 164–165.)

Ihde also probed beyond visualism. Multisensory technologies are commonplace in entertainment, such as cinema. Simulated learning environments, such as pilot simulators, were an early instance of virtual multisensory reality. In such applications, isomorphism with non-mediated reality remains important to re-create the situation in question. Virtual reality can be combined with isomorphic displays in augmented reality. Remote sensing and manipulation have become possible with robotics and teleoperation methods with feedback. Pilots of fighters and ships can be assisted by displays that allow visualizing large amounts of information in a single picture and make the environment observable even in difficult weather or underwater conditions. These technologies may not be science, but nothing prevents applying them in science as well, at least in principle. (Ihde 1998, 184–195.)

For the purposes of this article, I define *visualism* as the thesis that interpretation of the visual outputs of instruments is hermeneutic activity.

4. The Pragmatic Maxim

Charles S. Peirce introduced pragmatism in collaboration with William James (1842–1910) as a method for clarifying meanings. Peirce had already worked out some of its basic ideas by 1868–9 (see Peirce CP 5.213–357; EP 1, 11–82), but he published its first extensive formulation in his *Illustrations of the Logic of Science* of 1877–8, a series of articles (Peirce CP 2.619–660, 2.669–693, 5.358–410, 6.395–427; EP 1, 109–199). In the second article, “How to Make Our Ideas Clear” of 1878, he announced his thesis, which has traditionally been called “the pragmatic maxim”:

Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object. (Peirce CP 5.402), (Peirce EP 1, 132.)

He explained what he meant in two preceding passages:

[...] what a thing means is simply what habits it involves (Peirce CP 5.400; EP 1, 131).

Thus, we come down to what is tangible and practical, as the root of every real distinction of thought, no matter how subtle [*sic*] it may be; and there is no distinction of meaning so fine as to consist in anything but a possible difference of practice (Peirce CP 5.400; EP 1, 131).

As such, his thesis remains problematic. For instance, Horkheimer ([1947] 2004, 33) and others have pointed out that the pragmatic maxim leads to an infinite *progressus*, if taken literally. The conception of the object is the conception of its practical bearings; then, the conception of the practical bearings is the practical bearings of these practical bearings, and so on, ad infinitum. That may be problematic: if such progress arises, we have to wait infinitely long until we can determine the meaning of a thing. That is a serious problem and deserves a careful solution. I reply to this criticism at the end of this section.

Peirce himself added many times after 1900 that the practical bearings involve not only what actually takes place but also *conditional*, *potential* and *conceivable* effects – what is *discoverable* experimentally (Peirce CP 5.18, 5.196, 5.412, 5.425–427, 5.438, 5.453, 5.465; EP 2, 134–135, 145, 234–235, 332, 339–341, 346, 354, 356, 400–401). He adds in “What Pragmatism Is” in 1905 that *rational* meaning consists in *general kinds of experimental phenomena*, where “phenomenon” is to be understood objectively, like “effect” in “the Hall effect” or “the Zeeman effect.”⁶ He meant that *if* the actions of the experimenter fulfilled certain conditions, certain effects *would* follow. He also said that everything that influences conduct can be experimentally discovered. (Peirce CP 5.425–427; EP 2, 339–341.) This is what he means when he states that the pragmatist method of ascertaining the meaning of a symbol is the experimental method (Peirce CP 5.465; EP 2, 400–401). He also said in 1908 that nothing can be understood save in the terms of habits (Peirce CP 6.481; EP 2, 447–448). Short (2007, 173) called that revision “the subjunctive version of pragmatism.”

Peirce’s definition, as such, does not restrict meaning to language. *Anything* can be meaningful, provided that it has potential practical effects. On the other hand, Peirce himself *did* restrict the application of the pragmatic maxim to scientific and philosophical concepts (see e.g. Peirce CP 5.438, 5.467; EP 2, 346, 401–402), (Short 2007, 263), (Vehkavaara 2007, 268), (Stjernfelt 2014, 295). I do not see any reason for that; the maxim seems applicable to pretty much anything.

If I am correct, then Peirce’s maxim is fit for handling *non-linguistic meanings*. Though the term might look like an oxymoron to somebody trained in 20th century philosophy with its obsession about language (cf. Ihde 1998, 116), it does denote a class of phenomena. It includes aposematism, perceptual and motor functions of

⁶ Hence “phenomena” in Peirce’s sense are repeatable configurations of the *world*, not of a “subject.” This notion differs from the subjectivistic interpretation of phenomena of, for instance, Kant ([1781/7] 1956) and Niiniluoto (1999, 91). Hacking ([1983] 2010, 220–232) and Rouse (2002, 263–300) also have a notion of “phenomenon” which is strikingly similar to Peirce’s. I consider them synonymous.

cognition (Lakoff and Johnson 1999) and, as I will show in the next section, cause-effect relations conceived as means-ends relations. There seems to be no reason, why such phenomena could not be *described* in language. If there is no vocabulary, people are at liberty to introduce one. But that does not *make* them language.

To the best of my knowledge, not many people have tried to defend Peirce's maxim. Hence I restate a sketch of an argument published in Lindholm (2023). What is to be shown is that *the meaning of a thing consists in the potential practical effects of that thing*.

First, let us show that potential practical effects are a necessary condition of meaning. For a *reductio ad absurdum*, suppose that a thing has meaning but no potential practical effects. In the absence of potential practical effects, its existence would be completely insignificant for any embodied interpreter. This insignificance contradicts the hypothesis that the thing has meaning. In fact, we would not even know that such a thing existed in the first place. Hence having potential practical effects is a necessary condition of meaning.

Second, let us show that potential practical effects are a sufficient condition of meaning. For a *reductio*, suppose that a thing has potential practical effects but no meaning. This amounts to saying that a thing could potentially determine the outcome of a practical affair but we would not understand it. But taking such potential determination of an outcome into account already bestows the thing in question with meaning. Again, this contradicts the hypothesis. Hence having potential practical effects is also a sufficient condition of meaning.

Together these arguments show that meaning and potential practical effects are equivalent. Or at least so a Peircean could argue. Peirce himself devised different proofs for his maxim: an argument from phenomenology (Peirce CP 1.317–321; EP 2, 360–370), an argument from normative science (Peirce CP 1.573–574, 5.448n, 5.549–554; EP 2, 371–397), and an argument from semiotics (Peirce CP 1.560–562, 5.11–15, 5.464–496; EP 2, 398–433; MS 318). Peirce's arguments are significantly longer and more sophisticated than my rough-and-ready sketch.

Now I present my reply to Horkheimer's criticism which I have already given in Lindholm (2023). Peirce's semiotics, to which he dedicated his final years, shows how to solve this problem. However, his conceptual apparatus is complex; so I will simply sketch the essentials here.⁷

A *sign-relation* is a triadic relation between a *sign-vehicle*,⁸ an *object*, and a possible *interpretant*. The sign-vehicle can be *anything* that means something else to a potential interpreter – that is, it has potential practical effects. The interpretant is what the sign in its significant function essentially determines in its interpreter, or, in other words, the total proper effect of the sign taken by itself. (Peirce CP 2.242, 2.274; EP 2, 13, 272–273, 290–291, 409–410, 429.) There are different kinds of interpretants, and they constitute the meaning category. These definitions are *functionalist* (Stjernfelt 2014, 107): *anything* that functions like a sign-vehicle is a sign-vehicle; and this functionalism carries over to Peirce's subdivisions.

⁷ An interested reader may consult the second volume of *The Essential Peirce* (Peirce EP 2) and *Semiotics and Signifcans* (Peirce and Welby SS).

⁸ Sometimes Peirce calls the sign-vehicle *representamen*.

Peirce (CP 4.536, 8.183, 8.341; EP 2, 407, 480–481, 495–496, 498) distinguished between two objects: the *immediate object*, which is the object as the sign-vehicle represents it, and the *dynamical object*, which is the object independent of representation. The two objects are not like Kant's ([1781/7] 1956) phenomena and noumena; the dynamical object is completely knowable. Hence the distinction between the two objects is merely relative. The immediate object is part of the dynamical object and indicates, how the rest of the dynamical object can be known by *collateral experience* (Peirce CP 6.318, 6.338, 8.178–179, 8.183, 8.314; EP 2, 404–409, 429, 480, 493–498).

Peirce allowed the interpretant to be another sign-vehicle with the same dynamical object. In this process, the immediate object develops. Then the new sign-vehicle can elicit another interpretant. Thus there may arise an indefinitely long procession of signs interpreting previous signs – that is, *semiosis*. But semiosis is goal-oriented (Short 2007, 91–150, 158, 171–174), (Vehkavaara 2007, 263–264, 273): it tends towards improvement, and hence it will eventually terminate, if the dynamical object is completely known. The accumulation of knowledge about the dynamical object consists in the establishment of habits for dealing with it. The terminus can be a “quality of feeling,” an “exertion” (i.e. an action), or a habit of action. (Peirce CP 4.536, 8.332; Peirce MS 318.) Hence interpretation need not be intellectual; otherwise we could never get from thought to action (Short 2007, 201). In the terminus of semiosis, the immediate object coincides with the dynamical object, i.e. the dynamical object is completely known in all its aspects, and all possibilities for its interpretation, i.e. habitual encounters with it, have been exhausted.

By considering Peirce's semiotics, a *progressus* – semiosis – is indeed possible; but, ideally, it will terminate. By successive interpretations, the immediate object changes. If inquiry is ideal, this change will consist of gradual improvements. It is goal-oriented: it terminates at the discovery of the dynamical object – when the immediate object coincides with the dynamical object. Thereby a *final logical interpretant* – a habit of action – is formed. In this way, Peirce's semiotics is a description of the self-normative character of scientific inquiry. Peirce is not saying that real inquiry lives up to these norms. It is possible that real inquiry will never succeed in discovering the dynamical object; and even if it did, we might not know that. He is only saying that the norms that apply to the ideal case also apply to the real case, and the former provides a standard for the evaluation of the latter.

Thus Horkheimer was indeed correct in that, at least potentially, there arises a *progressus*, but it tends to termination. Semiosis terminates when the dynamical object is completely known. But pragmatists usually subscribe to Peirce's fallibilism,⁹ or the doctrine that anything can be questioned if positive grounds for doubt arise.¹⁰ If fallibilism is indeed true, then we have no criterion to determine when something is completely known and no further improvement in knowledge is possible.

⁹ For Peirce's own account of fallibilism, see “Questions Concerning Certain Faculties Claimed for Man” (Peirce CP 5.213–263; EP 1, 11–27); “Some Consequences of Four Incapacities” (Peirce CP 5.264–317; EP 1, 28–55); “The Fixation of Belief” (Peirce CP 5.358–387; EP 1, 109–123); and “How to Make Our Ideas Clear” (Peirce CP 5.388–410; EP 1, 124–141).

¹⁰ Isaac Levi (1983) is one of the notable exceptions. He rejected fallibilism in favour of what he called corrigibilism. I am unsure whether these positions differ except in name.

Moreover, the dynamical object can change, which makes a constant revision of all concepts necessary. Hence the final conclusion about anything must be postponed indefinitely (cf. Short 2007, 331). Thus Horkheimer's objection ceases to be an objection and becomes a description of the increase of knowledge in an idealized inquiry.

Fallibilism applies to pragmatism itself. Hence the argument sketched above should not be taken to claim universality but, rather, as a *hypothesis to be tested*. Pragmatists should acknowledge the possibility that their ideas can fail, even if they are widely applicable. Robin (1997) may have hit upon the truth when he suggested, in effect, that *the real proof of pragmatism is to apply it*. William James seems to have suggested a proof of pragmatism by living by it (Rydenfelt 2009, 48). The validity of pragmatism can be only established experimentally on strictly pragmatist grounds. Experimentation is needed to determine whether and how pragmatism fails. Then the possible negative results can be used to *improve* pragmatism. Hence pragmatism is akin to a *research programme* rather than a ready-made doctrine.

5. The Genesis of Meaning

John Dewey (1929b, 81–84) provided an account of how meaning emerges from experiment. First, a known change (possibly nothing) is introduced. Then something else (possibly nothing) changes as a result. This change is measured. Then these changes are correlated. The operations are repeated in varied conditions. If the correlation persists in different circumstances, an experimental practice can arise, which associates these changes. Thereby they become signs of each other: the presence of one is a (fallible) sign of the presence of the other.

The emergence of meaning consists in the reconceptualization of the discovered cause–effect relations as means–ends relations. Once a causal relation has been learned, by directly manipulating the occurrence of the cause, one acquires the capacity to indirectly manipulate the occurrence of the effect. Thus Dewey's account on the emergence of meaning is a direct consequence of Peirce's pragmatic maxim: the meaning of a cause or a means is its potential effects or ends. According to Dewey, that is the purpose of all intelligent activities. If meaning emerges from them, they acquire the status of *art*. (Dewey [1925] 1929a, 136, 177, 180–183, 369–370), (cf. Lindholm 2021.)

This discussion suggests one way in which the pragmatic maxim is inadequate. I have explained that Dewey seems to have maintained that the cause and the effect are signs of each other. But the pragmatic maxim says that only the cause can be a sign of the effect, not *vice versa*. I see no reason why that should be the case. For instance, the reading of a barometer provides information related to an effect of air pressure. But that makes the former a sign of the latter; that is the very purpose for which the instrument has been designed. Hence it seems that, at least in some cases, an effect can be a sign of its cause. It is strange that this did not occur to Peirce given that he studied similar examples of indices, a subclass of signs (see, e.g., Peirce CP 2.285, 5.73; EP 2, 8, 163). Hence I conclude that meanings *can* include *both* possible practical effects *and* possible practical causes. In short, to be meaningful is to be

potentially causally active.¹¹ At any rate, Peirce was on the right track. His crucial insight was that causality is constitutive of meaning. But he considered only one direction of causality – what will be possible in the future. Obviously, there is another direction – what may have been in the past. That is not a refutation but a generalization – a continuation of Peirce’s work.

Note that here the notion of “experiment” and the associated theory of the emergence of meaning are not confined to science in the ordinary sense. Dewey (1916a, 163–178, 237, 317–322) equated *all* experience with experiment at least once. He believed that experiment is the way infants learn their way through the world, anticipating Jean Piaget’s ([1936] 1952) concept of “little scientists.” All our beliefs are hypotheses that are being tested all the time, and we formulate new hypotheses when novel experiences prompt us. Here the notion of “hypothesis” is not confined to explicit theoretical statements but also denotes an implicit *plan of action* that is available to any animal independently of whether it uses language or not. Thus Määttänen (2009; 2015) has argued that the object of experience is not an assemblage of things or perceptions but an array of *opportunities for action*, or *affordances* (cf. Gibson 1979).

On the other hand, Dewey (1929b, 79–80, 84–85, 124, 199, 220, 240–242, 271, 295); (1938, 60–80) suggested an enlarged notion of science that would encompass any skilled activity that includes intelligent, critical insight for the improvement of its practices. He believed that science and common sense form a continuum of explicitness, publicity, precision and criticism, and that there is only a difference in degree, not in quality. Thus, in this sense, science is art and art is practice (Dewey [1925] 1929a, 354–393), (Lindholm 2021). Then the notion of “experiment” and the theory of the emergence of meaning *are* confined to science in Dewey’s enlarged sense.

Dewey’s account on the emergence of meaning can be directly applied to Rouse’s practical hermeneutics and Ihde’s visualism.

Practical hermeneutics is the thesis that practices are meaningful in their own right; that a practice consists of a configuration of people, their skills, their activities, their purposes, equipment, and environment; that interpretation is itself a practice; and that the world is interpreted by how we practically respond to it. Now, a practice is based on causal relations: people engage with it because it is a means towards an end in view. That applies primarily to the control of bodily movements and secondarily to the use of instruments external to the body.¹² A practice is meaningful because it is causally active. Its meaning has thus arisen from the discovery of its causal functions. Bodily movements become meaningful when one discovers what

¹¹ Causal activity could, perhaps, also be taken as the definition of *existence*. Then it must be conceptually distinguished from *reality*, that is, being mind-independent. See, e.g., (Peirce CP 1.324, 1.328–329, 1.457, 2.84, 3.200n3, 3.613, 6.336, 8.262).

¹² This distinction between primary and secondary causal functions is just a way to restate Dewey’s (1922, 24–26) distinction between active and passive means. Instruments in the ordinary sense are passive means; they acquire the status of a means only when actively used; outside their use they are just objects. On the other hand, habits are active means that have the power to activate a passive means. I have suggested in Lindholm (2021) that Rouse’s (1987, 70–126, 209–247); (1996, 125–259); (2002, 161–360) notion of “practice” might capture better what Dewey tried to convey with his notion of “habit.”

can be achieved with them. Instruments become meaningful when one discovers what can be achieved with certain bodily movements combined with these instruments.¹³

Visualism is the thesis that the interpretation of the visual outputs of instruments is hermeneutic activity. Now, by definition, what makes them outputs is that they are causally produced by such instruments. Hence the outputs are the meanings of these instruments. What makes them meaningful is how they are used in purposeful activities. Now, one of the purposes of science is to determine facts about objects. Some of them can be determined simply by naked-eye observation (provided that the observer is located and oriented appropriately). But many of them cannot; this requires the technical extension of the possibilities for observation. Instrumentation must ultimately transform some signals produced by the objects into signals observable to the naked eye, however many intermediate transformations that may take. That, in turn, requires knowledge about causal relations: how – that is, under what conditions – a signal can be transformed into another form. The purpose of the instruments is to provide such conditions. Each such transformation creates meaning in Dewey's sense: the signal to be transformed and the instrument that transforms it are the means; the signal that has been transformed is the end; the means and ends function as (fallible) signs of each other; and meaning is thereby constituted by their being causally related.

6. Technics as Hermeneutics

One can easily see that Dewey's account on the emergence of meaning is thoroughly *technical*. Technics is the activity of discovering and using means for achieving an end in view. This, in turn, is based on causal relations. Now, technics can be understood in at least two senses: as a set of skills or as a set of instruments. Practical hermeneutics accounts for both, and visualism seems to be its special case.

The universal character of practical hermeneutics resurrects the medieval notion of the *book of nature*, though in a secular sense. It can be understood as a neo-Kantian thesis, with some important differences. Neo-Kantians are correct in that all experience is interpreted, but they are wrong in considering all interpretation linguistic. (Cf. Määttänen 2009; 2015) I have argued for the existence of non-linguistic meanings. These allow direct access to the world. Hence universal hermeneutics does not necessarily make language an impenetrable barrier between us and the world. That notion dispenses with the sceptical problems that arise from representationalism. If all access to the world were mediated by language (as in theoretical hermeneutics), it would make sense to ask, whether there is any “world” at all beyond language (cf. Rorty 1997, 17). That places important restrictions on the Duhem–Quine thesis: that interpretation presupposes background *practice*, which

¹³ An instrument achieves nothing by itself unless it is an automaton. Whatever a non-automatic device achieves is a result from its being used with skilled movements of the body. Arguably, an automatic device, perhaps one with an adaptive control algorithm (Sastry and Bodson 1989; Åström and Wittenmark 1995), may be considered as having achievements of its own. See the previous note.

may or may not involve background theory. Such practices are mainly non-linguistic. They cannot be arbitrary, because they are based on discovered causal relations turned into means–ends relations – that is, technics. If there were no non-arbitrary non-linguistic meanings, then all interpretation would be arbitrary. Then, in principle, anything could be done with anything, which is hardly plausible.

Peirce's semiotics adds more conceptual resources to practical hermeneutics and visualism. I am not able to do justice to its complexity in this article;¹⁴ hence I confine myself to indicating how it allows direct access to the world. Peirce's best-known subdivision of sign-vehicles is that between *icons*, *indices*, and *symbols*. Icons signify their objects by resemblance (e.g. isomorphism). Indices signify their objects by being actually (e.g. causally) connected with them. Symbols are general signs that signify their objects by a rule, often a human convention. Most linguistic meanings can be analysed as symbols. This division provides a new insight into visualism. Visual displays are icons in the sense that they resemble some aspect of a phenomenon. Simultaneously they are indices by being causally connected, via technical intermediaries, with these phenomena. This makes technics an important anti-sceptical agency.¹⁵ Moreover, visual displays can be symbols insofar as they capture the general features of phenomena. The objects of such symbols are then multiply realizable, repeatable patterns or configurations of the world present in such phenomena.

Peirce also created an original theory of propositions on the basis of his semiotics, which Stjernfelt (2014) insightfully outlined. It is based on two other subdivisions of sign-vehicles: (1) that between *qualisigns*, *sinsigns* and *legisigns* (or *tones*, *tokens* and *types*), and (2) that between *rhemes*, *dicisigns* and *arguments*. Propositions are symbols, legisigns (or types), and dicisigns. They contain an indexical part, or a subject, that denotes a dynamical object, and an iconic part, or a predicate, that describes it. By that virtue, they bear truth-value. Propositions, being general types, are in some sense independent of their tokens, or the media of their expression. Hence many, if not all, propositions can be expressed by non-linguistic means. Sometimes nature itself, including ourselves, readily forms propositions for us.¹⁶ This provides a direct access to the world for us. But sometimes we have to create technical means – for instance, visual outputs like photographs, videos, etc. – to complement and supplement nature. In this sense, Ihde's visualism often, if not always, provides directly accessed propositional content about the world, independently of language.

¹⁴ Hence I advise again the interested reader to consult the first volume of *The Essential Peirce (Peirce EP 2) and Semiotics and Significality (Peirce and Welby SS)*.

¹⁵ This notion appears also in Marxism—Leninism. Kuusinen (1959, 98–99) dispels doubt by appealing to practice, technology and industry. Marx (MEW 3, 7) argued in the 8th thesis on Feuerbach that theoretical disputes are resolved by practice and the comprehension of that practice. Engels (MEW 21, 276–277) argued that our ability to produce a phenomenon ourselves confirms that our hypothesis about the phenomenon is at least approximately true. Thus “practice is the criterion of truth” (Kuusinen 1959, 111). Marx did not justify the 8th thesis on Feuerbach himself, but Dewey's notion of language as conceived means–ends relations can fill in for him. Dewey himself argued that we understand a phenomenon insofar as we are able to produce, sustain, and terminate it at will ([1925] 1929a: 428).

¹⁶ The whole fabric of biosemiotics is proof (cf. e.g. Barbieri 2007).

7. Criticism

Traditional philosophers of science might be upset with the thesis that natural science is a matter of interpretation. Somebody might have the prejudice that interpretation should be arbitrary. Hence, if that were correct, then the “knowledge” which science allegedly acquires would have no normative standing. Any belief or statement would be as good as any other. Essentially, anything would go (cf. Feyerabend [1975] 2010).

But, in fact, hermeneuticity need not threaten the objectivity of science. It is not the case that just anything goes. A single object may have several interpretations, but some of them will be better, some worse. Whether an interpretation is apt can be determined with reference to the object itself about which it is made and to the purpose for which it is made. Whether the purpose is attained is an objective matter of success or failure, which, in turn, depends objectively on the properties of the object and its relations to its environment. In this way, the object itself, the purpose of interpretation and the environment do *constrain* interpretation, but that does not necessarily mean *uniquely determining* it.¹⁷

For instance, a bicycle, cultivating vegetables, or the duck–rabbit (cf. Wittgenstein [1953] 2009, 204–209, 216–218) might be interpretable in a number of ways. A bicycle can indeed be taken as a muscle- or electricity-powered means of transportation along the surface of the Earth or an article to be sold; cultivating vegetables as producing food or as a means of relaxation; and the duck–rabbit as a duck or a rabbit. But it does not seem that somebody could plausibly interpret a bicycle as a spacecraft; cultivating vegetables as playing rugby; or the duck–rabbit as a portrait of J. V. Snellman. Perhaps that could be done satirically, for child’s play, or for some other such activity for which correctness of representation can be largely or completely irrelevant; but such use is derivative.

Not only does that hold for natural sciences but for human and social sciences as well. In any science, one must consider some evidence. The rejection of certain evidence can be acceptable but must be well justified. One cannot make arbitrary claims. For instance, judging whether religion *A* equals religion *B*, or whether François Rabelais and Nikolai Gogol have interesting stylistic similarities, must be based on, say, a comparison of the religious practices of the believers, or a stylistic analysis of the works of the authors. Again, satire or child’s play might be exceptions.

Another possible criticism turns on historical relativity. Technics changes. Hence our access to phenomena changes too. For instance, the methods of measuring gravity have changed drastically over centuries. That makes technically mediated knowledge historically relative, at least to a degree. When people abandon certain practices for several centuries, they may become unable to be resurrected. This is the case when a certain piece of equipment falls out of use, and after a period of time, nobody might know, why or how that instrument had been used. Hence some knowledge of the past can be lost forever.

¹⁷ These remarks elaborate on Ihde’s (1998, 197–198) anticipation of the same criticism.

I will bite the bullet here and concede that there is some room for historicism. That was already Kuhn's ([1962] 1996) and Rouse's (1987, 26–69) point. Instruments change. People use different instruments in different practices. Hence practices change too. Now, according to the pragmatic maxim, these different practices will have different meanings. Hence concepts and theories are incommensurable between different historical epochs. That may not be a problem as such, but it summons the spectre of relativism: without a universal measure of correctness, it becomes impossible to determine, whether science is progressing or not.

This kind of historicism may be troublesome for representationalist epistemology that conceives truth as accurate representation and progress as improvement in accuracy. But that is not the case with a pragmatic epistemology that identifies knowledge with practice or the ability to solve problems (cf. Dewey 1916b; [1925] 1929a, 354–393; 1929b; 1938); (Lindholm 2021). Let us suppose that our solution to a certain problem changes over time. We can still speak about progress if the change consists in *increasing the efficiency* of the whole body of possible solutions. We can also speak about progress if the change consists in *increasing the effectiveness* of the whole body of possible solutions; that is, if we are able to solve *more* problems than previously. Now the notion of the accuracy of representation may retain some relevance but it ceases to be the measure of goodness. Accuracy does matter in many, if not all, problems. But it is an empirical matter, *how* accurate is accurate enough in order to solve a given problem. A perfectly accurate representation is just a copy of the world, with all its indefinitely minute detail. Re-creating the situation would merely duplicate the problem. What could possibly be the point of such activity? Perhaps an immaterial Cartesian soul could be happy with this, but living organisms cannot: we have purposes and only so much resources. Logic is not self-sufficient but is a part of an ongoing activity: the conclusion of an inference is not the conclusion of an inquiry (Dewey [1910] 1933, 100–101); the solution must also be acted out to resolve the situation. Even if desirable, it is just practically impossible to re-create all the detail of a situation. Even if possible, it would be just useless to add detail when it no longer contributes to solving the problem. The relevant standards are *efficiency* (how much resources must be sacrificed to solve a problem) and *effectiveness* (the number of possible solutions); and the accuracy of representation has relevance only insofar as it contributes to either.

Dewey did not dispense with the notion of truth, however. He subscribed to the correspondence theory of truth. He just emphasized that the application of the truth–predicate is confined to the domain of possible experience (Dewey 1941, 178–179) and that it is not strictly distinct from utility (Dewey 1916b, 240–241, 324–325).

The body of possible solutions of problems is composed of causal knowledge; and if meaning *is* such knowledge, it constitutes the meaning of the world. We understand the world by coping with it. With each solution, our life becomes more meaningful and significant. (Dewey [1910] 1933, 18–21; 1929b, 124–135.)

Dewey's theory of language (1916a, 17–19; [1925] 1929a, 166–207; 1929b, 140–169; 1938, 42–59) may deserve some criticism. He has two notions of language: language as *conceived means–ends relations* and language as *communication*. These two senses overlap partially, but are at least conceptually distinct. It is not always clear,

which definition he is using. The first sense includes a large number of meanings that are clearly non-linguistic – even though Dewey calls them “language.” On the other hand, Peirce (EP 2, 221) was clear that his semiotics includes, but does not reduce to, language.

8. Conclusion

I have argued that if the pragmatic maxim is sound, then the pragmatist theory of meaning provides a theoretical basis for Rouse’s practical hermeneutics and Ihde’s visualism. I have also suggested that this account can – and indeed should – be supplemented by Peirce’s semiotics.

I have defended Peirce’s pragmatic maxim against Horkheimer’s criticism. Peirce’s semiotics turns it into a description of semiosis. I have also suggested that the maxim is a hypothesis to be tested.

I have cited Dewey’s account on the emergence of meaning from experiment. That seems to provide a counterexample to the pragmatic maxim, but I have turned it into its enlargement. I have shown how to derive Rouse’s practical hermeneutics and Ihde’s visualism from that thesis of Dewey.

I have also defended universal hermeneutics, based on the pragmatist theory of meaning, against two possible criticisms that stem from the relativity of interpretation and the historical relativity of practices.

My account has been only preliminary so far. But it is easy to expand it and elaborate on it. I have already alluded to the conceptual apparatus of Peirce’s semiotics. The analysis of scientific practices and instruments can be continued in semiotic terms. In this way, followers of Rouse and Ihde can learn from classical pragmatism.

On the other hand, the classical pragmatists say relatively little about practices in detail. I have pointed out in Lindholm (2021) that even Peirce, who practised experimental science himself, keeps details at arm’s length. Rouse’s and Ihde’s analyses of particular practices and instruments can fill in the missing detail. Hence the classical pragmatists can also learn from Rouse and Ihde.

One possible line of inquiry that this discussion suggests is the re-evaluation of naïve realism – that is, the position that the world is as it is experienced. At least since Kant ([1781/7] 1956), philosophers have distinguished between subjective phenomena and objective noumena. He reformulated Parmenides’s influential distinction between appearance and reality (Diels and Kranz 1960, 28A, B) in modern terms. But Dewey (1929b, 295) rejected both distinctions: he argued that the world as it is experienced *is* the real world; the predicate “real” makes sense only within the confines of experience.¹⁸ Likewise, Rouse (1987, 7–8) argued that the way things

¹⁸ Here it must be borne in mind that classical pragmatists did not consider experience as exclusively “subjective.” Dewey was most emphatic to define *experience* as *biological and social life* (1916a, 2, 6–7) and as organism—environment interaction (Dewey 1916a, 163–178, 37, 317–22; 1938, 1–98; 1941, 183–184). In this sense, experience is a public, observable, and causal process and hence merits the status of objectivity, whatever “subjective” traits it may have besides. Arguably, this notion had appeared already in Peirce (CP 1.324, 1.336).

appear in inquiry is their reality. If Dewey and Rouse are right, then “phenomena” do not denote “subjective” events but, rather, repeatable patterns or configurations of the *world*¹⁹ in the way I have already suggested. Arguably, a case can be made for naïve realism in these terms. Dewey himself did so at least once (1916b, 250–263), but his arguments may have dated and hence might need revision.

Another possible line of inquiry is a comparison of the pragmatist theory of meaning with the received view in analytic philosophy that propositions are *abstract entities*.²⁰ In light of Dewey’s account on the genesis of meaning, it seems readily clear that he conceives all meanings – including propositions – as *constructs*. For Dewey (1916b; 1938), propositions are tentative and hypothetical plans of action, the purpose of which is to solve a problematic situation. Hence they are thoroughly *technical*. If Dewey is correct, they are real but hardly abstract and certainly not entities. They do not exist independently of their potential application. They seem to be better captured by adjectives, adverbs, verbs or gerunds that characterize possible courses of action than nouns (Dewey [1925] 1929a, 158–9; [1934] 1980, 263); (Hickman [1990] 1992: 10). This non-reifying view is at odds with analytic philosophy but seems to have received some support from second-generation cognitive science (Lakoff and Johnson 1999).

References

- Åström, Karl Johan, and Björn Wittenmark. *Adaptive Control*. Reading, MA: Addison-Wesley, 1995.
- Barbieri, Marcello (Ed.). *Introduction to Biosemiotics: The New Biological Synthesis*. Cham: Springer, 2007.
- Dewey, John. *The Influence of Darwin on Philosophy, And Other Essays in Contemporary Thought*. New York: Henry Holt & Co, 1910.
- Dewey, John. *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. (Revised edition.) Lexington, MA: D. C. Heath and Company, [1910] 1933.
- Dewey, John. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: The Macmillan Company, 1916a.
- Dewey, John. *Essays in Experimental Logic*. New York: Dover, 1916b.

¹⁹ Here “phenomenon” must be understood as synonymous to “effect,” as in “the Hall effect” or “the Zeeman effect” (Peirce CP 5.425–427; EP 2, 339–341).

²⁰ In the mainstream, propositions are usually conceived reductively: either as unstructured sets of possible worlds in which they are true; or as structures composed of universals and particulars. Some take propositions to be *sui generis* or a class of entities not reducible to anything else. There are also others, who understand propositions (1) as *facts* composed of individuals, properties and relations (Jeffrey C. King), (2) as *properties* of being such that the subject instantiates the predicate (Jeff Speaks), or (3) as cognitive event types of *predication* (Scott Soames) (cf. King, Soames and Speaks 2014).

-
- Dewey, John. *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Henry Holt and Company, 1922.
- Dewey, John. *Experience and Nature*. (Revised edition.) London: George Allen & Unwin, Ltd, [1925] 1929a.
- Dewey, John. *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York: Minton, Balch & Company, 1929b.
- Dewey, John *Art as Experience*. New York: Perigee, [1934] 1980.
- Dewey, John. *Logic: The Theory of Inquiry*. New York: Henry Holt and Company, 1938.
- Dewey, John. "Propositions, Warranted Assertibility, and Truth." *The Journal of Philosophy* 38, no. 7 (1941): 169–186.
- Diels, Hermann and Walther Kranz *Die Fragmente der Vorsokratiker; Griechisch und Deutsch. Erster Band*. (Neunte Auflage.) Berlin: Weidmannsche Verlagsbuchhandlung, 1960.
- Dilthey, Wilhelm. *Gesammelte Schriften, Band V*. Stuttgart: Teubner, 1956.
- Dilthey, Wilhelm. *Gesammelte Schriften, Band VII*. Stuttgart: Teubner, 1957.
- Duhem, Pierre. *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press, [1906] 1954.
- Feyerabend, Paul. *Against Method: Outline of an Anarchistic Theory of Knowledge*. New York: Verso, [1975] 2010.
- Gibson, James J. *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum, 1979.
- Hacking, Ian. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press, [1983] 2010.
- Heidegger, Martin. *Sein und Zeit (Gesamtausgabe, Band 2)*. (Friedrich-Wilhelm von Herrmann, Ed.) Frankfurt am Main: Vittorio Klostermann, [1927] 1977.
- Hesse, Mary. *Revolutions and Reconstructions in the Philosophy of Science*. Bloomington: Indiana University Press, 1980.
- Hickman, Larry A. *John Dewey's Pragmatic Technology*. Bloomington and Indianapolis: Indiana University Press, [1990] 1992.
- Horkheimer, Max. *Eclipse of Reason*. London and New York: Continuum, [1947] 2004.
- Husserl, Edmund. *Krisis der europäischen Wissenschaften und die transzendente Phänomenologie: Eine Einleitung in die phänomenologische Philosophie, herausgegeben von Walter Biemel (Husserliana Band VI. 2. Auflage)*. (Walter Biemel, Ed.) Haag: Martinus Nijhoff, [1936] 1976.
- Ihde, Don. "The Experience of Technology'." *Cultural Hermeneutics* 2 (1974): 267–279.
- Ihde, Don. *Experimental Phenomenology: Multistabilities*. (2nd expanded edition.) Albany: SUNY Press, [1977] 2012.
- Ihde, Don. *Technics and Praxis: A Philosophy of Technology*. Dordrecht, Boston and London: D. Reidel, 1979.
- Ihde, Don. *Technology and the Lifeworld: From Garden to Earth*. Bloomington and Indianapolis: Indiana University Press, 1990.
- Ihde, Don. *Expanding Hermeneutics: Visualism in Science*. Evanston, IL: Northwestern University Press, 1998.
- Kant, Immanuel. *Kritik der reinen Vernunft*. Hamburg: Felix Meiner, [1781/7] 1956.
- King, Jeffrey C., Scott Soames and Jeff Speaks. *New Thinking about Propositions*. Oxford: Oxford University Press, 2014.

- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. (3rd edition.) Chicago: University of Chicago Press, [1962] 1996.
- Kuusinen, Otto Wille (Ed.). *Osnovy marksizma–leninizma. Uchebnoe posobie. (The Fundamentals of Marxism–Leninism. A textbook.)* Moskva: Gosudarstvennoe izdatelstvo politicheskoy literatury, 1959.
- Lakoff, George, and Mark Johnson. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books, 1999.
- Latour, Bruno. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press, 1987.
- Levi, Isaac. “Truth, fallibility and the growth of knowledge.” In *Language, Logic and Method*, edited by Robert S. Cohen and Marx W. Wartofsky: 153–174. Dordrecht: Reidel, 1983.
- Lindholm, Juho. “Science as Art: A Deweyan Perspective.” In *The Yearbook of the Irish Philosophical Society 2020–2021. Special Issue: Science and Politics* (Cara Nine, Ed.), 2021. Accessed March 14, 2023.
http://www.irish-philosophical-society.ie/wp-content/uploads/2021/06/Lindholm_Science-asArt_IPS-2021-Final.pdf
- Lindholm, Juho. “The Pragmatist Basis of Material Hermeneutics”, in *World Academy of Science, Engineering and Technology International Journal of Humanities and Social Sciences* 16, no. 11, 2022.
<https://publications.waset.org/10012795/the-pragmatist-basis-of-material-hermeneutics>
 [accessed August 17, 2023]
- Lindholm, Juho. “Scientific Practices as Social Knowledge.” In *International Studies in the Philosophy of Science*, 2023. Accessed July 19, 2023.
<https://doi.org/10.1080/02698595.2023.2196930>
- Marx, Karl and Friedrich Engels. *Karl Marx Friedrich Engels Werke* (MEW). (46 volumes.) Berlin: Dietz Verlag, 1956—.
- Määttänen, Pentti. *Toiminta ja kokemus: Pragmatistista terveen järjen filosofiaa (Action and Experience: Pragmatist Common Sense Philosophy)*. Helsinki: Gaudeamus, 2009.
- Määttänen, Pentti. *Mind in Action: Experience and Embodied Cognition in Pragmatism*. Cham, Heidelberg, New York, Dordrecht and London: Springer, 2015.
- Niiniluoto, Ilkka. *Critical Scientific Realism*. Oxford and New York: Oxford University Press, 1999.
- Peirce, Charles Sanders. *The Collected Papers of Charles Sanders Peirce* (CP). (8 volumes. Hartsthorne, Charles, Paul Weiss and Arthur W. Burks, Eds.) Cambridge, MA: Harvard University Press, 1958.
- Peirce, Charles Sanders. *The Essential Peirce: Selected Philosophical Writings* (EP). (2 volumes. Houser, Nathan, and Christian J. W. Kloesel and The Peirce Edition Project, Eds.) Bloomington: Indiana University Press, 1998.
- Peirce, Charles Sanders. *Manuscripts* (MS). In *Charles Sanders Peirce Papers* (MS Am 1632). Cambridge, MA: Houghton Library, Harvard University. (1966: *The Charles S. Peirce Papers, Microfilm Edition, Thirty Reels with Two Supplementary Reels Later Added*. Cambridge, MA: Harvard University Library Photographic Service. Catalogued in Robin, Richard S. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. Amherst: University of Massachusetts Press. Also available online at
<http://hollis.harvard.edu/?itemid=1000602456> and at
<https://id.lib.harvard.edu/alma/990006024560203941/catalog>

-
- Peirce, Charles Sanders, and Victoria Welby. *Semiotic and Significs: The Correspondence between C. S. Peirce and Victoria Lady Welby* (SS). (2nd edition. C. S. Hardwick & J. Cook, Eds.) Bloomington and Indianapolis: Indiana University Press, 2001.
- Piaget, Jean. *The Origins of Intelligence in Children* (Vol. 8). New York: International University Press, [1936] 1952.
- Polanyi, Michael. *Personal Knowledge: Towards a Post-Critical Philosophy*. London: Routledge, [1958] 1962.
- Polanyi, Michael. *The Tacit Dimension*. Gloucester, MA: Peter Smith, [1966] 1983.
- Quine, Willard van Orman. "Two Dogmas of Empiricism." In *The Philosophical Review* 60 (1951): 20–43.
- Robin, Richard S. "Classical Pragmatism and Pragmatism's Proof." In *The Rule of Reason: The Philosophy of Charles Sanders Peirce*, edited by Jacqueline Brunning and Paul Forster: 139–152. Toronto: University of Toronto Press, 1997.
- Rorty, Richard. *Truth, Politics, and 'Post-Modernism'*. Assen: Van Gorcum, 1997.
- Rouse, Joseph. *Knowledge and Power: Toward a Political Philosophy of Science*. Ithaca and London: Cornell University Press, 1987.
- Rouse, Joseph. *Engaging Science: How to Understand Its Practices Philosophically*. Ithaca and London: Cornell University Press, 1996.
- Rouse, Joseph. *How Scientific Practices Matter: Reclaiming Philosophical Naturalism*. Chicago and London: The University of Chicago Press, 2002.
- Rydenfelt, Henrik "Pragmatism and the Aims of Inquiry: Peirce, James and Two Seeds of Death." In *Pragmatist Perspectives (Acta Philosophica Fennica, Vol. 86)*, edited by Sami Pihlström and Henrik Rydenfelt: 41–51. Helsinki: HakaPaino, 2009.
- Sastry, Shankar, and Marc Bodson. *Adaptive Control: Stability, Convergence and Robustness*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- Short, Thomas Lloyd. *Peirce's Theory of Signs*. Cambridge: Cambridge University Press, 2007.
- Stjernfelt, Frederik. *Natural Propositions: The Actuality of Peirce's Doctrine of Dicisigns*. Boston: Docent Press, 2014.
- Vehkavaara, Tommi. "From the Logic of Science to the Logic of the Living. The Relevance of Charles Peirce to Biosemiotics." In *Introduction to Biosemiotics: The New Biological Synthesis*, edited by Marcello Barbieri: 257–298. Cham: Springer, 2007.
- Wittgenstein, Ludwig. *Philosophical Investigations*. (4th revised edition. Anscombe, Gertrude Elizabeth Margaret, Peter Michael Stephan Hacker, and Joachim Schulte, Eds.) Chichester: Wiley-Blackwell, [1953] 2009.
- Wittgenstein, Ludwig. *The Blue and the Brown Books*. Oxford: Blackwell, 1958.
- Wittgenstein, Ludwig. *On Certainty*. (Anscombe, Gertrude Elizabeth Margaret, and Georg Henrik von Wright, Eds.) New York: Harper and Row, 1969.

The Falsificationist View of Machine Learning

Machine learning pushes the frontiers of algorithmic achievements, though the striving for state-of-the-art performance often obscures the fragility of enforcing decisions amid uncertainty. This paper interprets machine learning within Karl Popper's epistemology. We assess machine learning paradigms' fit for falsificationism and argue that the new interpretation can improve robustness. Though the price is to accept unambiguous decisions, the restriction of the hypothesis space still adds value. The context for our work is established by comparison with similar techniques and highlighting its limitations.

Keywords: *machine learning, epistemology, artificial intelligence, falsificationism, Popper, robustness*

Acknowledgment

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) and ELLIS for supporting Patrik Reizinger.

Author Information

Patrik Reizinger, University of Tübingen, IMPRS-IS, ELLIS

<https://orcid.org/0000-0001-9861-0293>

How to cite this article:

Reizinger, Patrik. "The Falsificationist View of Machine Learning".

Információs Társadalom XXIII, no. 2 (2023): 117–127.

== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.7> ==

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

Falsificationism (Popper 2002) provides a sobering view of scientific progress—an insight generally neglected by engineers. Applying modern scientific advancements requires making decisions in complex environments, but optimizing performance often sacrifices robustness.

Nassim Nicolas Taleb (2020, 2007) argues that the 21st century challenges humanity with Black Swans—highly improbable events with considerable losses. Such events are the reality of solutions in medicine (Monti, Zhang and Hyvärinen 2020) or autonomous driving (Szemenyei and Reizinger 2019) – often powered by artificial intelligence (AI). Although researchers made notable progress in protecting neural networks against adversarial examples (Schott et al. 2018) and quantifying uncertainty (Gawlikowski et al. 2021), the authors argue that the field could benefit from adopting Popper’s philosophy. That is, the process of falsification: a careful evaluation of neural networks’ predictions.

The belief of obtaining reliable, task-specific models with a limited amount of data and the ever-increasing pressure to achieve state-of-the-art performance obscure the fragility of the quest for perfection in a noisy setting: the need for a decision disregards whether the best solution is reliable and superior compared to the alternatives, resulting in notorious failures.

The Popperian flavor of mathematical methods is not unknown: statistical hypothesis testing. (Gretton et al. 2007) provides conclusions based on falsifying hypotheses, and is applied, e.g., in software testing (Tóth et al. 2017). This paper examines modern machine learning methods in the falsificationist context, arguing that constraining the hypothesis space would improve decision quality and reliability, as opposed to striving for an unambiguous decision.

Popper’s philosophy inspired several researchers in the sciences, even in the broad context of learning systems. Berkson and Wettersten (1984) showed that falsificationism can be seen as learning theory—nonetheless, differences were also highlighted, e.g., by comparing Popper’s degree of falsifiability and the Vapnik-Chervonenkis (VC) dimension of statistical learning theory (Corfield, Schölkopf and Vapnik 2009). Vasconcelos, Cardonha and Gonçalves (2018) utilized Popper’s philosophy for fair hiring decisions.

Similar to Vasconcelos, Cardonha and Gonçalves (2018), we rely on falsificationism and utilize it in a broad sense. Namely, Popper’s claim—that probabilistic statements are neither verifiable nor falsifiable (Popper 2010)—would render his arguments invalid for probabilistic machine learning. However, defining a decision threshold for probabilities admits a falsificationist view of probabilistic systems. By drawing parallels to Popper’s philosophy, our goal is to motivate an epistemological view of machine learning.

Although we start from a theoretical assessment, our conclusions focus on the practical. We claim that falsifying hypotheses can potentially reduce false predictions and express uncertainty while providing additional information compared to traditional machine learning approaches. According to our best knowledge, this is the first work that interprets modern machine learning in a Popperian way.

First, we discuss supervised, self-supervised, unsupervised, and reinforcement learning. We conclude that the Popperian approach is generally applicable. Since unsupervised learning lacks hypotheses, falsificationism can only apply if we have a priori assumptions about the representation. Our analysis contrasts classification and regression methods—pointing out that falsificationism naturally fits only the former, though techniques exist for the latter as well.

Second, we compare the merits of falsificationism to other robustness-improving strategies, such as predicting confidence or using ensembles. Third, we address the case of adversarial examples to point out the limitations of our proposal. By providing an epistemological context for machine learning algorithms, we hope to inspire a discussion that improves the robustness of AI.

2. The falsificationist machine learning taxonomy

This section analyzes machine learning concepts from a falsificationist point of view, assesses their suitability for the paradigm, and also addresses practical implications.

2.1. Terminology

Machine learning algorithms constitute a mapping between different domains: the data fed to the network is called the input sample or observation, and the output of the network is the prediction (this can be discrete or continuous). When discrete, it is generally called a label/class, whereas prediction or action can be both. In the discrete case, we speak of classification, in the continuous, of regression. The network learns and uses a (latent) representation, whereas the goal specified by the designer is prescribed by the objective/loss.

2.2. Supervised Learning

Supervised learning learns a mapping given pairs of inputs and desired outputs. Examples include image classification (the image is the input and the label is the desired output) or stock price prediction (a time series is the input; the next element in the future is the desired output). The difference compared to rule-based algorithms is that, although we know the desired output, we cannot specify how to produce it from the inputs.

Classification

Standard algorithms maximize the correct class's probability—incorrect labels can have arbitrarily close probabilities unless they are less than the correct one. The falsificationist interpretation requires that all incorrect labels' probabilities are pushed toward zero—this is equivalent to maximizing the correct label's probab-

ity. To contrast the differences, consider classification with three labels. Technically, the correct class is predicted as soon as it has the highest probability. However, this approach is agnostic to the difference to the second-highest probability: it considers the solution correct even if the three probabilities are 0.331, 0.33, and 0.299 (the correct class having the highest probability). Minimizing the incorrect classes' probabilities requires the highest possible difference, implying a notion of robustness.

This is similar to the loss (called hinge loss) of support vector machines (SVMs) (Schölkopf et al. 1999), where the model is incentivized to increase the correct label's probability above the others plus a margin. As the margin goes toward one, we recover the falsificationist approach.

Pushing the label distribution to a Dirac delta is not novel—compare it with the review of Gawlikowski et al. (2021)—though its falsificationist interpretation is. We advocate for the falsificationist approach for a more reliable prediction and a safer failure. Namely, hard-to-classify samples could potentially have a predicted label distribution with multiple large entries. Driving any of those to zero—even if not leading to a definitive conclusion—can help to reduce the hypothesis class yielding (partial) suspension of judgment, since the output is a label set, but it is reduced.

Does this mean that the model does not provide added value? On the contrary, as the hypothesis space is reduced, the false certainty of a crisp decision is alleviated. Acknowledging the practical limitations, we should not deceive ourselves that algorithms can always make a good decision.

A similar approach to the falsificationist view in classification is set-valued prediction, where a model generally outputs the labels with the k highest probabilities (Lapin, Hein and Schiele 2016; Mortier, Hüllermeier and Waegeman 2022). This approach reduces the hypothesis class in the same way as the falsificationist approach—one could also predict the labels above a specific probability threshold.

Regression

Regression differs from classification in having an infinite label set. Theoretically, this is no burden to adapt the falsificationist view on the predictions (possible theories in the sciences are also uncountable), but it restricts practical applicability, as excluding infeasible options will not lead to the correct solution. To overcome this problem, one could divide the values into mutually exclusive categories—turning the regression problem into classification—but that would sacrifice resolution.

2.3. *Unsupervised Learning*

Unsupervised algorithms—e.g., clustering algorithms such as k -Means or t -SNE (Van der Maaten and Hinton 2008)—utilize unlabeled information to extract a “good” representation that can be used to solve multiple downstream tasks—their

competitive advantage is avoiding the expensive labeling process. Since the desired output is unknown, the only feedback about performance is via the objective function.

Architectures such as Variational AutoEncoders (VAEs) (Kingma and Welling 2013) learn a representation and a generative model for, e.g., generating realistic images—i.e., there are no labels but a continuous reconstruction loss with a Kullback-Leibler divergence as inductive bias. Thus, a falsificationist interpretation does not apply to the predictions. We could treat models with an error above a threshold as falsified, but this would not bring us closer to the solution—it would reason about the reconstruction, not the representation.

Moreover, reconstruction quality is only a necessary indicator of a high-quality representation. As Alemi et al. (2018) point out, even a meaningless latent representation can produce good samples.

Thus, falsificationism is not practically applicable on the predictions, for we cannot restrict the hypothesis space, similar to regression—lacking the desired output, we do not even know the hypothesis space. However, if we assume that the underlying representation has particular (measurable and testable) properties such as independence, then evaluating such properties can be a basis for falsification. But the clear difference compared to any other paradigms is that the falsificationist view concerns the properties of the representation, not the predictions.

2.4. *Self-Supervised Learning*

Self-supervised learning (Lil’Log 2019) resembles supervised learning in the sense that labels are predicted from observations, except that the labels are generated by the model itself. This auxiliary labeling process can exploit unlabeled data (similar to unsupervised learning) and makes falsification feasible in some scenarios.

2.4.1. *Generative Adversarial Networks (GANs)*

GANs (Goodfellow et al. 2014) consist of a discriminator and a generator. The discriminator distinguishes between real and generated (“fake”) images, whereas the generator’s role is to deceive the discriminator by producing realistic samples. The architecture casts generative modeling into binary classification, so falsificationism is applicable. Nonetheless, this bears no practical advantage as excluding one label means the same as accepting the other.

2.4.2. *Contrastive Learning*

Contrastive learning (Lil’Log 2021a; Chen et al. 2020) is the perfect practical example of falsificationism. It learns a representation via an auxiliary classification task by combining samples with different (negative pairs) and same labels (positive pairs). Its objective incentivizes alignment, uniformity, and separability (Wang and

Isola 2020). By alignment, samples of the same class are forced to have a similar representation, whereas negative pairs are mapped to different latents, ensuring separability. Uniformity incentivizes evenly distributed representations in the latent space.

Positive and negative pairs reflect how scientists verify hypotheses. If the representation (“the hypothesis”) is not able to match specific observations (the positive pairs), then it is necessarily wrong. Nevertheless, this is not sufficient; we can imagine that if all samples get mapped to the same representation (known as mode collapse), then the alignment is perfect, but the representation is meaningless. By enforcing different representations for negative pairs, they are “repelled” from each other, rendering them distinguishable.

The similarity is more evident via hard negative mining (Robinson et al. 2020), which collects negative samples with similar representations but different labels. Hard negative samples contain more information, so they help refine the decision boundary between classes and improve separability. Thus, hard negatives are more informative and have a higher degree of falsifiability.

2.5. *Semi-supervised Learning*

Semi-supervised learning combines supervised and unsupervised strategies (Lil’Log 2021b) and (Lil’Log 2021a), relying on a large unlabeled and a much smaller labeled dataset. Some methods utilize both datasets simultaneously, whereas others rely on unsupervised data for representation learning (pre-training) and then deploy the labeled samples for fine-tuning.

Being a mixture of two paradigms, we can rely on the conclusions of the parts. Generally, supervised learning admits the falsificationist approach on the prediction, whereas unsupervised learning does not—since in this case the representation is used for a supervised task, it might not be meaningful to discuss the falsifiability of the representation’s properties. Depending on classification/regression, the same considerations apply as above.

2.6. *Reinforcement Learning*

Reinforcement learning resembles how humans learn: a decision-making agent interacts with its environment via its actions and receives a reward (feedback). As the agent does not have access to the optimal policy, which it aims to learn, it can compare two actions only relatively, based on the received reward.

Actions can be discrete or continuous, so the classification and regression arguments—discussed for supervised learning—apply. The only difference is that here we do not have the correct output. Nonetheless, we can utilize the same practical strategies to restrict the possible solutions (e.g., when predicting an action from a discrete set); thus, bearing the benefits of the falsificationist approach.

3. Extensions

This section discusses strategies to increase the robustness of machine learning models. We assess how using “unknown” labels, confidence scores, uncertainty estimates, and ensembles relate to falsificationist machine learning.

3.1. “Unknown” labels

Including a label for objects from unknown classes (usually denoted as UNK) in classification can signal uncertain decisions (Radford et al. 2019)—and makes the hypothesis space complete. This strategy is crucial in out-of-distribution (OOD) data (Yang et al. 2021), or the extreme case of changing environments, where not only the class probabilities change but also new classes are introduced.

One could argue that the UNK label expresses suspension of judgment. We believe this is partially true, since it can express that with a single prediction the output is inconclusive. However, it can fail for hard-to-categorize data. Imagine an animal classification task with cows, -cats, dogs, and UNK. Assume the prediction assigns 0.501 to the “dog” and 0.499 to the “cat” label. Despite including UNK, a standard algorithm would predict “dog,” although UNK expresses the uncertainty better. However, predicting UNK would neglect the information that the hypothesis space is restricted to dogs and cats (based on the probabilities of this example). The falsificationist view would suspend judgment and give the reduced label set “dog” and “cat.”

Still, it is beneficial to include the UNK label. Assume the same probability distribution as before, but the object is a lion. Not including UNK would imply the object is a dog or a cat. Suspension of judgment is still the correct answer, but the hypothesis space of “dog” and “cat” would be incorrect. This example highlights that falsificationist machine learning and the UNK label are orthogonal: falsificationism makes the decisions more robust (even with sacrificing unambiguity) for a *given* hypothesis space, whereas the UNK label extends the hypothesis space to account for distributional changes.

3.2. Confidence scores

Confidence scores express prediction feasibility with a value in $[0;1]$. In object detection (Szemenyei and Reizinger 2019), when the number of objects can vary within scenes, confidence represents whether an object is present; thus, filtering out false positives is a compensation for the architectural bias that fixes the number of objects. In natural language applications—such as in Microsoft (2021)—their purpose is to estimate chatbot answer quality. We focus on the latter case.

Is a high confidence score necessary or sufficient for correct predictions? Assume an image of an animal resembling both a dog and a cat (which is not unrealistic). Our classifier predicts the label and its confidence and assigns a probability of 0.501 to “dog” and 0.499 to “cat.” Assume that the prediction is correct; the image contains

a dog. Low confidence would be consistent with the probabilities (the same holds when the image is of a cat), though it would not contain additional value compared to the label distribution. Although high confidence would not contradict the correctness of the prediction, it could not express the label distribution’s uncertainty.

Thus, we conclude that confidence scores can be redundant to the label distribution for prediction quality assessment. Moreover, predicting confidence—similar to the UNK label—neglects information: a “dog” label with a confidence of 0.6 is less informative than outputting the label set “dog” and “cat” and requires an additional mechanism (to predict the confidence scores), whereas the falsificationist approach does not.

On the other hand, when confidence describes another property (e.g., the presence of an object as in the object detection case), it remains useful, as such information is not contained in the labels. That is, if no object is present, predicting zero confidence scores for all classes could be a potential solution; however, including a separate class label of no object might also be sufficient.

3.3. *Ensembles*

Ensembles (also known as expert systems) pool different models to increase prediction performance and reliability (Tifrea, Stavarache and Yang 2021; Pathak, Gandhi and Gupta 2019; Masegosa 2020) with multiple strategies, e.g., consensus or majority vote for classification and weighting for regression.

If the models agree, their prediction is accepted. In the case of classification, this means that all/most models predict the same label, whereas regression needs a tolerance (e.g., to assess whether the predictions are “close enough”). On the other hand, the prediction is falsified when one model draws a different conclusion. Even in that case, the set/range of predictions for classification/regression could be used as the reduced hypothesis space—a clear practical advantage.

Ensembles can potentially assess representation quality in unsupervised learning (what we want to extract) instead of proxies, such as reconstruction quality. Even beyond testing properties of the representation (such as independence), it might also be possible that an ensemble can identify models with good representation but poor reconstruction through raising a flag if all ensemble members report equivalent representations but possibly suboptimal reconstruction quality—such as VAEs with a good encoder but a poor decoder (Alemi et al. 2018). However, this requires a unique representation (or a correction for invariances) since generally it is not guaranteed that even the same neural network learns the same representation when trained multiple times.

3.4. *Uncertainty estimates*

Probabilistic machine learning quantifies the prediction’s uncertainty via probability distributions’ variance/entropy instead of single-point estimates (Gawlikowski

et al. 2021). We acknowledge that these methods improve robustness and argue for their use in the falsificationist framework. Instead of “only” assessing uncertainty, the available information can also restrict the hypothesis space. For example, the variance in regression can describe a feasible interval.

4. Limitations

Although falsificationism can improve the robustness of machine learning by interpreting the predictions differently (perhaps with a change of objective function but the same architecture), it is not the Holy Grail.

We see falsificationism as a means of improving predictions’ reliability near the decision boundary—where multiple options are possible. On the other hand, adversarial examples—samples malevolently modified by exploiting the networks’ computational properties to trick the model into believing that the sample belongs to a different class—will have a negative effect, unless a defense is put into place. This is irrespective of using the falsificationist approach or not. Namely, such examples exploit the model’s properties, and as falsificationism does not change the architecture but mainly the interpretation of the predictions, it has no means to defend against adversarial attacks automatically.

5. Conclusion

This paper draws inspiration from Karl Popper’s falsificationist philosophy and investigates its applicability to the predictions of machine learning models. We contrasted theoretical and practical arguments and reinterpreted existing machine learning methods. Our main conclusion is that—mainly by interpreting the predictions differently—it is generally possible to adapt the falsificationist view.

Moreover, we voiced our support for doing so, as the deployment of machine learning in critical systems requires robustness—we believe that starting to think with the falsificationist mindset could help achieve this goal.

References

- Alemi, Alexander, Ben Poole, Ian Fischer, Joshua Dillon, Rif A. Saurous, and Kevin Murphy. “Fixing a broken ELBO.” In *35th International Conference on Machine Learning vol. 80*, 159–168. Stockholmsmässan, Stockholm Sweden: PMLR, 2018.
- Berkson, William, and John Wettersten. “Learning from Error, Karl Popper’s Psychology of Learning.” *Synthese* 78, no. 3, 1989.
- Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. “A simple framework for contrastive learning of visual representations.” In *37th International Conference on Machine Learning vol. 119*, 1597–1607. Stockholmsmässan, Stockholm Sweden: PMLR, 2020.

-
- Corfield, David, Bernhard Schölkopf, and Vladimir Vapnik. “Falsificationism and statistical learning theory: Comparing the Popper and Vapnik-Chervonenkis dimensions.” *Journal for General Philosophy of Science* 40, 51-58, 2009.
- Gawlikowski, Jakob, Cedric Rouvère, Njéoutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler, and Xiao Xiang Zhu. “A survey of uncertainty in deep neural networks.” arXiv preprint arXiv:2107.03342, 2021.
<https://doi.org/10.48550/arXiv.2107.03342>
- Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets.” *Advances in neural information processing systems* 27, 2014.
- Gretton, Arthur, Kenji Fukumizu, Choon Teo, Le Song, Bernhard Schölkopf, and Alex Smola. “A kernel statistical test of independence.” *Advances in neural information processing systems* 20, 2007.
- Kingma, Diederik P., and Max Welling. “Auto-encoding variational bayes.” arXiv preprint arXiv:1312.6114, 2013.
<https://doi.org/10.48550/arXiv.1312.6114>
- Lapin, Maksim, Matthias Hein, and Bernt Schiele. “Loss functions for top-k error: Analysis and insights.” In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 1468-1477. Las Vegas, NV, USA: CVPR, 2016.
- Lil’Log. Weng, Lilian. “Contrastive Representation Learning.” 2021a. Accessed May 31, 2021.
<https://lilianweng.github.io/posts/2021-05-31-contrastive/>
- Lil’Log. Weng, Lilian. “Learning with Not Enough Data Part 1: Semi-Supervised Learning.” 2021b. Accessed December 5, 2021.
<https://lilianweng.github.io/posts/2021-12-05-semi-supervised/>
- Lil’Log. Weng, Lilian. “Self-Supervised Representation Learning.” 2019. Accessed November 10, 2019.
<https://lilianweng.github.io/posts/2019-11-10-self-supervised/>
- Masegosa, Andres. “Learning under model misspecification: Applications to variational and ensemble methods.” In *Advances in Neural Information Processing Systems* 33, 5479-5491. NeurIPS Virtual-Only Conference, 2020.
<https://proceedings.neurips.cc/paper/2020>
- Microsoft. “Confidence Score – QnA Maker – Azure Cognitive Services.” 2021. Accessed January 17, 2022.
<https://docs.microsoft.com/en-us/azure/cognitive-services/qnamaker/concepts/confidence-score>
- Monti, Ricardo Pio, Kun Zhang, and Aapo Hyvärinen. “Causal discovery with general non-linear relationships using non-linear ica.” In *35th Uncertainty in Artificial Intelligence Conference vol. 115*, 186-195. Tel Aviv, Israel: PMLR, 2020.
- Mortier, Thomas, Eyke Hüllermeier, Krzysztof Dembczyński, and Willem Waegeman. “Set-valued prediction in hierarchical classification with constrained representation complexity.” arXiv preprint arXiv:2203.06676, 2022.
<https://doi.org/10.48550/arXiv.2203.06676>
- Pathak, Deepak, Dhiraj Gandhi, and Abhinav Gupta. “Self-supervised exploration via disagreement.” In *36th International Conference on Machine Learning vol. 97*, 5062-5071. Long Beach, CA, USA: PMLR, 2019.

- Popper, Karl. *The Logic of Scientific Discovery*. London: Routledge, 2002.
<https://www.routledge.com/The-Logic-of-Scientific-Discovery/Popper/p/book/9780415278447>
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. “Language models are unsupervised multitask learners.” *OpenAI blog* 1, no. 8 (2019.)
- Robinson, Joshua, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. “Contrastive learning with hard negative samples.” arXiv preprint arXiv:2010.04592, 2020.
<https://doi.org/10.48550/arXiv.2010.04592>
- Schölkopf, Bernhard, Robert C. Williamson, Alex Smola, John Shawe-Taylor, and John Platt. “Support vector method for novelty detection.” *Advances in Neural Information Processing Systems*, 1999.
- Schott, Lukas, Jonas Rauber, Matthias Bethge, and Wieland Brendel. “Towards the first adversarially robust neural network model on MNIST.” arXiv preprint arXiv:1805.09190, 2018.
<https://doi.org/10.48550/arXiv.1805.09190>
- Szemenyei, Marton, and Patrik Reizinger. “Attention-based curiosity in multi-agent reinforcement learning environments.” In *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, 176–181. Athens, Greece: IEEE, 2019.
<https://doi.org/10.1109/ICCAIRO47923.2019>
- Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007.
- Taleb, Nassim Nicholas. “Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications.” arXiv preprint arXiv:2001.10488, 2020.
<https://doi.org/10.48550/arXiv.2001.10488>
- Țifrea, Alexandru, Eric Stavarache, and Fanny Yang. “Novelty detection using ensembles with regularized disagreement.” In *ICML 2021 Workshop on Uncertainty & Robustness in Deep Learning*, 2021.
<http://www.gatsby.ucl.ac.uk/~balaji/udl2021/accepted-papers/UDL2021-paper-100.pdf>
- Tóth, Tamás, Ákos Hajdu, András Vörös, Zoltán Micskei, and István Majzik. “Theta: a framework for abstraction refinement-based model checking.” In *2017 Formal Methods in Computer Aided Design (FMCAD)*, pp. 176-179. IEEE, 2017.
- Van der Maaten, Laurens, and Geoffrey Hinton. “Visualizing data using t-SNE.” *Journal of machine learning research* 9, no. 11, 2008.
- Vasconcelos, Marisa, Carlos Cardonha, and Bernardo Gonçalves. “Modeling epistemological principles for bias mitigation in AI systems: An illustration in hiring decisions.” In *AIES '18: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 323-329. New York: Association for Computing Machinery, 2018.
<https://doi.org/10.1145/3278721.3278751>
- Wang, Tongzhou, and Phillip Isola. “Understanding contrastive representation learning through alignment and uniformity on the hypersphere.” In *International Conference on Machine Learning*, pp. 9929-9939. PMLR, 2020.
- Yang, Jingkan, Kaiyang Zhou, Yixuan Li, and Ziwei Liu. “Generalized out-of-distribution detection: A survey.” arXiv preprint arXiv:2110.11334, 2021.
<https://doi.org/10.48550/arXiv.2110.11334>

Frameless Folk Psychology

How Online Context Collapse and User-Centeredness Distort Social Cognition and Nudge Hostility and Epistemic Injustice

In this paper, I will argue that the rise in hostility and polarization on social media is explainable by taking into account a radical difference between online and face-to-face interaction. In everyday offline environments, socially shared and context-dependent norms frame the understanding of other people's minds based on their behavior. I will argue that, on social media platforms, social cognition is distorted thanks to two deliberate design choices that are a means for financial gain for the platform's designers: namely, the lack of socially shared norms on these platforms (entailed by what is known as context collapse) and their interfaces' extreme user-centeredness. I will argue that such design features not only cause frustration in the understanding of others but encourage testimonial injustice in interaction.

Keywords: *online hostility, social networking sites, social cognition, epistemic injustice, reactive attitudes*

Acknowledgments

I want to thank Prof. Marc V. P. Slors, Dr. Selene Arfini, Prof. Evert van der Zweerde, Dr. Fleur Jongepier, and Antonia Leise for their feedback and suggestions throughout the production of this paper.

Author Information

Giacomo Figà Talamanca, Radboud University, Nijmegen
<https://orcid.org/0000-0002-1623-6785>

How to cite this article:

Talamanca, Giacomo Figà . "Frameless Folk Psychology".
Információs Társadalom XXIII, no. 2 (2023): 128–145.
 ══ <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.8> ══

*All materials
 published in this journal are licenced
 as CC-by-nc-nd 4.0*

1. Introduction¹

Toxicity, polarization, and hostility are *du jour* concepts in contemporary research on social networking sites (SNSs). Further, SNSs enable people to be exposed to and engage with a variety of information sources, as well as interacting with people holding beliefs and views different from their own (Beam et al. 2018; Flaxman, Goel, and Rao 2016). However, they are also linked to a variety of aggressive and toxic behaviors, such as flaming (Moor, Heuvelman, and Verleur 2010; Rost, Stahel, and Frey 2016), cyberbullying (Kowalski et al. 2014; Langos 2012), trolling (Bail et al. 2020; Phillips 2015), dehumanization (Pacilli et al. 2016; Harel, Jameson, and Maoz 2020), and incivility (Masullo Chen et al. 2019; Rains et al. 2017).

I propose an explanation of how polarized aggression arises on SNSs such as Facebook and Twitter, by underlining the role of socially shared, context-dependent norms that normally characterize social interaction and understanding in everyday, offline settings. I argue that hostility in these platforms arises due to a frustration caused by a lack of shared grounds for mutual understandability. Not only does such lack of grounds make hostility in online interaction more likely but it favors epistemic (and specifically testimonial) injustice. I argue that the tendency to hostility and epistemic injustice is caused by deliberate design choices that characterize these platforms, namely, their user-centeredness and what is known as context collapse.

In Section 2, I introduce the theory of social cognition called “the regulative view of folk psychology,” which stresses the importance of socially shared and context-dependent norms for the everyday understanding of other people’s minds. In Section 3, I describe how certain SNSs such as Facebook and Twitter are characterized by context collapse and argue that it causes, on these platforms, a lack of socially shared norms. In Section 4, I argue that context collapse and the extreme user-centeredness of these platforms are the main contributors to hostility in online interaction. Specifically, they hinder social cognition by depriving shared norms and values that can act as a normative frame for facilitating mutual understandability. In Section 5, I argue that, by the same design choices, these platforms play an active role when their users commit testimonial injustice toward other people online.

2. Regulative Folk Psychology: The Role of Norms in Social Cognition

Before introducing the regulative view of folk psychology, it is worth briefly sketching the meaning of the term folk psychology within contemporary philosophy of mind. The term was introduced by Sellars (1956) to propose the way “ordinary” folk (i.e., non-philosophers or psychologists) understand each other’s minds while engaged in social interaction. Up until the early 2000s, the predominant theories of how people understand each other’s minds were the *Theory Theory*, which views social cognition as reliant on people knowing what mental states are, how they relate

¹ An earlier version of this manuscript, with a different title and which was part of the author’s master’s thesis, is available at Radboud University’s Thesis Repository: <https://theses.uhn.ru.nl/handle/123456789/10933>.

to action and to each other (Nichols and Stich 2003); and *Simulation Theory*, which views mental state attribution as reliant on the capacity to model others' mental states in one's own mind (Gallese 2009; Goldman 2006). However, since the diffusion of phenomenology and 4E cognition approaches in philosophy of mind, many competitor theories of social cognition emerged, such as primary interaction theory (Gallagher 2001, 2004), folk psychology as a narrative practice (Hutto 2009), direct social perception (Zahavi 2001, 2011), participatory sense-making (De Jaegher and Di Paolo 2007; Torrance and Froese 2011), and pluralistic folk psychology (Andrews 2012, 2015).

I wish to point out that these approaches are competing: endorsing one of these approaches entails excluding the others. In contrast, the approach I am going to endorse here can be seen as complementary to all these approaches as it highlights an aspect of social cognition rather than aiming to capture the essence of social cognition. This approach, known as “mindshaping” (Mameli 2001; Zawidzki 2013) or “regulative folk psychology” (Castro 2020; McGeer 2015, 2021), focuses on the role played by context-dependent, socially shared norms in mental state attribution, and is compatible with other contemporary theories of social cognition (see, e.g., Peters 2019; Westra 2020).

McGeer (2015) illustrates the role of norms in social cognition by describing, as an intuition pump, playing chess as a paradigmatic case of social interaction. To be a competent chess player, correctly attributing mental states to the other player – i.e., understanding the other player's beliefs, intentions, and desires – is crucial. To acquire such competence, the player must first learn the rules of the game. Learning these rules also requires a degree of sensorimotor skills, to understand the movement patterns of each piece on the board. Relevantly, when you don't follow the rules of the game, the other player is entitled to correct you and compel you to act in accordance with the rules – and if you intend to be a chess player, you will also feel compelled to follow them. Not only is being able to understand the mental state of the other player essential to being a competent chess player; the capacity is *grounded* in the normative infrastructure of the game itself and our sensitivity to the rules that compose it.

Everyday social interaction and mental state attribution work likewise as in the game of chess.² When we attribute beliefs, intentions, and mental states to others, the context where the interaction takes place has a grounding role. Just as in the rules of chess, there are norms underlying the different contexts of social interaction that mediate the expectations and interpretations of others' behavior. In every social context where interaction takes place, there are norms and conventions of appropriate and inappropriate behavior: human beings are raised in situated social environments characterized by (spatially and temporally variable) norms, including evaluative standards, ideals, values, codes of conduct, and/or imperatives that are shared within that social context. In other words, the context has a normative influence not only on how people behave but also on the interpretation of people's

² With the difference that the normative structure of everyday interaction is not as clear-cut and straightforward as a game of chess.

behavior. Context-dependent norms ground our expectations for other people's and our own behavior.³

The importance of contextual norms for mutual understanding is grounded, as Zawidzki (2013) notes, in the centrality of cooperation in our species. Being able to make commitments with one another to achieve a goal, and to jointly see that goal as making us accountable toward one another – i.e., to *establish a common ground* between people acting together – is central to carrying out that goal successfully. Such capacity to establish a common ground is central to a hyper-social species like ours.⁴ For this reason, on the one hand the normative (or “binding,” if you will) aspects of the context are inevitably varied due to the varieties of joint action human beings commonly partake in. On the other hand, it is because joint action is so central to human everyday life that we are characterized by such sensitivity to norms and to the feedback we receive from other people when we (do not) adhere to them. The relevance of norms in understanding other people's minds derives from the fact that our everyday agency is grounded in our social nature and in the importance of cooperation from an evolutionary standpoint.

For McGeer (2015, 2019), the grounding of social cognition and interaction in socially shared norms is highlighted by the intrinsic connection between understanding other people and the (moral) evaluation of their behavior and character. She makes use of the idea, introduced by P. F. Strawson, of *reactive attitudes*, the “non-detached attitudes and reactions of people directly involved in transactions with each other ... of such things as gratitude, resentment, forgiveness, love, and hurt feelings” (Strawson 1962, 5). Strawson argues that these attitudes – reacting to others' behavior through appraisal, disapproval, resentment and so on – have the function to hold other people liable to certain normative standards regarding (morally) appropriate and inappropriate behavior. When we assume these attitudes, either we expect our target to understand and conform to normative standards or we are illustrating to our target what those standards are and, eventually, why they should conform to them. If folk psychology is a normative practice, and the kind of assessment present in reactive attitudes implies a normative evaluation of behavior, then reactive attitudes can be considered a way of structuring and directing the target's behavior to a correct standard. Conformity with context-dependent norms matters for successful and smooth social interaction to take place, and not only the felt need to correct inappropriate behavior but the sensitivity to these corrections is fundamental to

³ Importantly, embodied expressivity – including facial expressions, gaze, tone of voice – is also important in our sensitivity to social feedback and understandability (Gallagher 2001; Zahavi 2011). For the general goals of this paper, the importance of embodiment will not be considered: for while there is a relevant sense where embodied expressivity is strongly hindered online, it is not necessarily accurate to claim that human interaction online is *disembodied* (see Osler 2021 for a discussion).

⁴ I am not claiming that the capacity to make commitments is the only relevant aspect of joint action; nor is this the place to explore in depth the debate regarding the ontology of joint action (Bratman 2014; Gilbert 2013) or whether other animals are capable of normative practices or making commitments (Figà Talamanca 2021; Vincent, Ring, and Andrews 2018). I just want to state that, in the case at hand of members of the species *Homo Sapiens*, the capacity to establish and feel compelled by a common ground is particularly important.

understand others' behavior and to make oneself not just understandable but accountable on the basis of such understandability.⁵

3. Online Context Collapse: From Merged Audiences to Lack of Shared Norms

Let us now turn to describe one of the most salient differences between face-to-face and online social interaction in the case of several SNSs. In a sense, such difference is characteristic of computer-mediated communication as such: as Kiesler et al. (1984) point out, because the message recipient is not tangibly present,⁶ “communicators must imagine their audience, for at a terminal it almost seems as though the computer itself is the audience” (1125). In the specific case of many SNSs – most prominently Facebook and Twitter – this problem persists in a slightly different form. While on such SNSs there is arguably a clearer understanding that a user’s communicative acts are received by other people, SNS users do not technically know who is going to see what they post. The content they share and post on such platforms can reach unexpected or unintended recipients. Typically, an SNS user will have online social ties with people who belong to different social groups in face-to-face interactions: for instance, your family members, your friends, your colleagues are likely to be parts of your own *online* social network, while you would interact with them in distinctly separate settings in face-to-face interaction. Even strangers, under certain circumstances, can interact with the content you generate:⁷ users do not have full control over who is going to consume their content and how (Selinger and Hartzog 2016). A Twitter or Facebook user may not know who exactly is going to see and react to what they post, or how they will react, because the “audience” of their posts is undetermined, constituted by a merging of different social groups that in offline interaction would normally remain distinct. This dynamic, which is typical of many SNSs but can also occur in some specific face-to-face interactions – such as weddings and funerals – is known as *context collapse*: “the flattening out of multiple distinct audiences in one’s social network, such that people from different contexts become part of a singular group of message recipients” (Vitak 2012, 451).

This is typical of many Web-based platforms to an extent, and is a feature that users of those technologies adapt to, deal with, or even make use of in a variety of ways. It is a particularly researched mechanism in SNSs, primarily in terms of its implications for users’ privacy management (Marwick and boyd 2014; Vitak 2012) and

⁵ Importantly, the shared norms, values, and goals that ground our understandability and accountability can vary greatly, ranging from procedural norms (such as the rules of chess) to common interests, shared goals, and moral norms and values existing in a given society.

⁶ To think of computer-mediated communication as lacking *any* form of embodiment is not obvious (Osler 2021). In the context of SNS-based communication (in contrast to, e.g., a private chat or a videocall service) what matters is not the “disembodiment” of the other but their indeterminacy.

⁷ For instance, when you comment under a Facebook page, people who also follow it (or are recommended that post by the algorithm on their newsfeed) but are not a social tie on your online network may read your comment.

news sharing and exposure (Beam et al. 2018; Kim and Ihm 2020).⁸ However, with little exception (Fox and Moreland 2015), the impact of context collapse in online interaction has not been researched as much. I intend to propose, here, that context collapse is one of the most radical and disruptive differences between face-to-face and online interaction. Before I specify why that is, there are two important aspects of online context collapse to be pointed out: first, what makes context collapse in SNSs such as Facebook or Twitter specifically distinct from offline instances of context collapse; and second, why context collapse on SNSs is often a deliberate design choice.

To make the difference between online and offline context collapse clear, one can take a wedding party as an example of offline context collapse. Context collapse is the merging of normally separated social contexts. In a wedding, different people from different social contexts of your life can be present – your parents, your colleagues, your college friends, your conservative relatives. Each of them may have diverse expectations regarding how one should behave at a wedding and may tend to hold other participants accountable to their own standards: in fact, because of their diversity, these expectations may clash – and it may happen that, say, your conservative aunt reprimands your college friends for making inappropriate jokes. However, if such a scenario occurs, other people (including the newlyweds) are entitled to quash the fight in the name of having a nice wedding. In other words, while (some) expectations regarding appropriate behavior at a wedding may vary, there are norms that characterize (offline) social practices that are *not* contested, and that characterize the social practice *as such*. If a wedding party is the celebration of the union of two people, expectations about certain aspects of behavior in that context – such as how to dress, or what jokes to make – may not be defined, but others – such as a fight between guests – clearly contravene the core commitments of the participants to the wedding’s course and overall purpose. These norms participants commit to also work to *ameliorate* the eventual setbacks of offline context collapse – of which the squabble between your friends and your aunt is an example. As argued before, the norms underlying the diverse social contexts of human interaction exist to enable and facilitate joint action: the norms entitling the newlyweds to reprimand people misbehaving at a wedding are an example of this function. In cases of offline context collapse such as a wedding, there still exist norms that mediate interaction and frame the understanding and moral assessment of other people. On SNSs, as environments characterized by online context collapse, there are no such norms: the merging of different audiences online, which in offline environments is mediated by norms that can facilitate interaction, is not characterized by any clear or shared values, norms, or standards among those involved. This is a key difference

⁸ Note that much of current research on context collapse focuses on SNS use in the Western world, which can limit the understanding of these platforms’ social impact: for instance, Costa (2018) found that teenage Facebook users in Turkey create different profiles for each of their social contexts (family, friends, schoolmates, etc.) in order to avoid any awkwardness online that context collapse would generate. However, while creating multiple profiles can help in avoiding context collapse, such choice contravenes the platform’s intended use and underlying ideology from the standpoint of its designers, who want each user to create a singular profile (see below).

between offline and online context collapse: while instances of the former are characterized by norms that mediate and frame interaction and mutual understanding for the purposes of joint action, instances of the latter are not.⁹

The second aspect of online context collapse to be underlined is its role as a deliberate design choice for SNSs such as Facebook and Twitter. While not explicitly acknowledged as such, in the case of Facebook, the merging of different audiences within the indeterminacy of the platform can be tracked as an intentional choice in the notorious statement by Zuckerberg: “You have one identity ... The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly ... Having two identities for yourself is an example of a lack of integrity” (Kirkpatrick 2010, 199). It is safe to assume that context collapse is not implemented simply due to Zuckerberg’s philosophical beliefs on personal identity. Rather, context collapse can be seen as a means for the SNS’s service providers to maximize collection of users’ data, which is the designers’ main source of revenue (Büchi et al. 2019; Hildebrandt 2008; Hildebrandt and Gutwirth 2008). The SNSs enable their designers to collect data through user engagement and that data is then sold to third parties. It is, in other words, in the designers’ interest that the collected data reflect as much as possible each user’s profile, character, and preferences. Zuckerberg’s statement on identity, reflected by context collapse as a design choice, can then be seen in a new light when keeping in mind that data collection is a means for service providers to achieve their financial gain. The interest in what he calls “integrity” really reflects the desire to gain as much information as possible about each user’s character and preferences, which are not necessarily manifest in each offline context that person lives in – at least, not in all at the same time. Context collapse can be seen as a way to extract a user’s character, preferences, and desires, by placing the user in an (online) environment where all of the people they interact with in their offline life are present and, in a sense, by attempting to paint a “complete” picture of that person by detaching them from the specific contexts of family, workplace, friendships, etc. Context collapse can be seen as a way for SNSs to maximize user data collection, by placing the platform’s focus on individual users abstracted from their diverse social contexts. And while this (somewhat Cartesian) view, evident by these platforms’ design, of the human self as existing separately from its embodied and embedded life can be considered inaccurate or problematic on its own, I intend to examine its implications for users’ online interpersonal life.

4. Hostility as Frustrated Interaction and the Self-Centeredness of Social Media Platforms

We now get to the crux of my argument. Social cognition and interaction in everyday, face-to-face settings relies on socially shared norms and conventions that not

⁹ An alternative way to capture the difference is the following: while cases of offline context collapse exist to favor forms of joint action, the use of SNSs such as Facebook and Twitter is not a joint action: ironically, while we engage with other people on such sites, we are not acting *with* other people.

only drive people's behavior but frame the understanding of such behavior. On SNSs such as Facebook and Twitter there are no such norms: due to online context collapse, social interaction on SNSs can be characterized as decontextualized as there is no shared frame for the evaluability and understandability of people's behavior. In other words, the capacity to understand other people's minds is severely undermined thanks to the lack of a shared normative frame, which would normally ground mutual understanding – in many SNSs, social cognition is frameless.

The lack of a normative frame for understanding others does not in itself make it impossible to understand others' minds: rather, such lack leads to a severe disruption of such capacity. Such disruption of the frame for understanding others' behavior online is understandable once we look again at how SNSs are designed, and what kinds of action they make salient to their users. Building on the previous section, I will now argue that the grounds on which people interpret and evaluate others online are determined by their own normative standards, and that the platforms where the interactions occur are responsible for such a coordinate shift.

So how can people online, when deprived of a socially shared and (to some extent) non-contested normative frame for interaction, interpret other people's behavior? The very design of these SNSs' interfaces as choice architectures can provide an answer to this question. The designers' goal is data collection through user engagement: the more people engage with content on the SNS, the more data are gathered and can be monetized by the service providers.¹⁰ Context collapse, as a design choice, can be considered a means for gathering a "full" picture of a user's profile, with the assumption that a complete portrait of a user's self is to be found abstracted from the diverse and situated social contexts in which that user lives. However, this function of context collapse is effective especially because the interfaces of SNSs such as Facebook and Twitter are entirely focused on users themselves. Besides the mechanism of notifications to draw users to the platform when they are not using it, and the way the algorithms of such SNSs provide certain kinds of content based on previous user engagement (Burr, Cristianini, and Ladyman 2018), the very way a Facebook or Twitter homepage is designed, in line with the designers' financial interest, is entirely centered on the user's own preferences and thoughts. From the recommendation to users of content and social ties (friends and other people to follow); to the chance for users to pick and personalize their online social network and favored information sources with great ease; to the continuous proposal of the most high-ranked content in users' feed and of content produced by or interacted with by the user's social ties; to the very simple invitation in the middle of the interface promoting users to express themselves, epitomized by Facebook's tagline "What are you thinking?" or Twitter's "What's happening?" – each and every aspect of these SNSs is entirely focused on the individual user's beliefs, thoughts, events surrounding them, expectations, and values. Users are prompted, or nudged, to see the (online world) through their own lens, their own beliefs, moral standards, norms, and values, as this facilitates their engagement with and the overall business model of these plat-

¹⁰ Due to the focus of this paper, I will not explore here issues concerning the gamified and often addictive (Eyal 2014) design of these technologies, which are in themselves important means for maximizing user engagement.

forms. However, this emphasis on the user comes at the cost of values that can be shared among the collective of people using the platform. And if shared norms are lacking, the grounds that enable mutual understandability and evaluability between people in everyday life significantly hinder successful interaction.

As Rini (2017) points out, as the norms of communication of SNSs are disputed and there is no common understanding of why people share and generate content, users of SNSs find themselves in an epistemically uncertain environment. For Rini, taking a partisan stance toward seemingly immoral content – and, specifically, sticking to one’s own normative frame of reference, one’s own values and standards for (in)appropriate behavior – can be considered rational. Exactly because there is no undisputed reference frame, SNSs users may choose to refer to their own – an action that is all but discouraged by the platforms’ interface design. If anything, as Marin (2021) argues, setting aside a few explicit (and minimally enforced) norms in the Terms and Conditions of such platforms, the only (somewhat implicit) meta-norm that these SNSs implement to guide users’ online life is a *norm of sociality*¹¹ promoting the expansion of people’s online social network. It is true that, under this general norm, users can create many social practices, or “language games,” with diverse goals and values. However, belonging to this group requires minimal conditions (such as following the same page, person, or information source at best, or being accepted by a group moderator after a request to join at worst). Due to context collapse as encountered by a user toward an imagined and somewhat all-encompassing community of users (in contrast to members of a specific niche, or followers of a specific page), the possibility of contact, attunement, or shared ground among the performers of such diverse language games is not encouraged by the SNSs at best and is incredibly uncertain at worst.

So here is the crux of social cognition on online platforms affected by context collapse: the understanding and evaluation (both moral and epistemic) of other people normally relies on socially shared values and norms, but due to online context collapse and the interface design, what frames the understanding of other people’s minds is the individual’s own beliefs, normative standards, and values, *which are not necessarily shared by other people online*.¹² Not only through a lack of shared conceptual resources (i.e., a lack of shared norms, standards, and values) but through being actively encouraged by the interface design and SNS algorithms, people tend to use their own standards, norms, and values when interpreting the behavior of

¹¹ She specifically speaks of a “meta-norm” of sociality, referring to Horne (2001).

¹² In the case of platforms such as Facebook and Twitter, this can be particularly highlighted by Marin’s (2021) notion of *general social media platform*. While what she calls *purposeful social media platforms* (such as dating apps, job sites, or Q&A sites) are joined by people with a clear and shared goal, and who (and whose generated content) can be flagged and excluded when openly contravening that goal and the shared standards necessary to achieve it, general social media platforms do not have a clear shared purpose or goal, as people can use them for a variety of different reasons and with different goals. The latter platforms include those that are characterized by context collapse, and their use is (almost) completely left up to individuals. On purposeful social media platforms, the presence of shared rules and goals entails a common ground that can enable people to act together and functionally interact with one another; on general social media platforms characterized by context collapse, users are not acting together – using a general social media platform does not constitute an instance of joint action, which requires shared norms, goals, and mutual understandability.

others, even when those norms do not correctly apply. The combination of context collapse and user-centeredness prompts people to rely on their own frame for orienting themselves in the understanding and (moral) evaluation of other people's behavior, even if the person whose behavior is interpreted follows a completely different normative frame for evaluating their own behavior and that of others. It is the lack of a shared normative frame that radically distorts social cognition on SNSs.

Ironically, while SNSs such as Facebook or Twitter seem focused on the social connections of individual users and on the transmission and sharing of content within a user's social network, these platforms are focused on the user alone. Information regarding individuals' connections and belonging to social groups is instrumental to maximizing engagement from individual users. For this purpose, these platforms subtly promote users seeing the world through the lens of *their own* values and expectations, a type of experience that is nudged both by the available choice of architecture of the platform (promoting users to share their own thoughts, notifications, and recommendations based on previous engagement) and by the absence of a shared social structure entailed by context collapse.

From an individual's view, the absence of a unified frame has the implication of (mis)representing behavior that follows different and unseen normative frames as morally and epistemically incompetent and will lead users to treat others as such. Our reactive attitudes toward those pieces of behavior, our attempts to expect and enforce conformity in others to a set of understandable and shared normative frame are not just unjustified and ineffective due to the absence of a shared frame. Likely, the other person will feel called out without justification, and consider your behavior immoral and unjustified; consequently, they will react to your behavior by assessing it as immoral and assessing you as the morally and epistemically incompetent agent. The application of each other's frames is derived from lack of a common ground, leading to interaction becoming dysfunctional and frustrated. Online hostility becomes common in the absence of a shared frame enabling mutual understandability and joint action. Not only, then, are social cognition and interaction frustrated due to the design choice of context collapse; SNSs such as Facebook and Twitter exploit our natural sensitivity to norms (both following and enforcing them) for financial gain, as frustrated cognition and reactive attitudes represent a source of engagement, regardless of the discomfort experienced due to the lack of a normally shared normative frame and regardless of the groundlessness of people's reactive attitudes.¹³

The empirical evidence available seems to confirm such a relation between unease, frustration, and outrage, on one hand, and sensitivity to (contested) normative expectations in SNSs, on the other.¹⁴ The observed tendency to either mediate self-pres-

¹³ This idea, while seemingly derogatory, is not unrealistic. In September 2021, as part of a vast document leak from Meta, the *Wall Street Journal* (Hagey and Horwitz 2021) reported that Facebook posts with a higher number of "angry" reactions were promoted by the algorithms much more than posts with a higher amount of "likes" or "love" reactions.

¹⁴ While the following overview of empirical studies retroactively motivates the elaboration of my argument, it might be relevant to think briefly of ways to verify it through empirical research. For instance, one might examine cases of conflictual online interaction by observing whether language related to norm enforcement is used, and whether such cases of conflictual interaction are more frequent when people from different online social networks interact.

entation in a collapsed context (Marwick and boyd 2011; Vitak et al. 2015; Wang and Mark 2017) or outright avoid context collapse through multiple profiles (Costa 2018) can be seen as a way to avoid any awkwardness or ambiguity in online interaction. While in the former case users negotiate their decontextualized self to interact without conflict, in the latter users avoid discomfort entirely by creating separate contexts for each profile they have. Fox and Moreland (2015) also observed that Facebook users had varying expectations regarding what kind of content is appropriate to share both on the SNS as such and with specific social ties, thinking that there are supposed to be “natural boundaries” for appropriate behavior while acknowledging that such standards can vary from person to person, as the platform does not establish or enforce any. Specifically: “[A]lthough participants report[ed] engaging in *rule development* for themselves, they often avoid the process of *boundary coordination* with others in the Facebook sphere. Instead, they choose to simply avoid offenders’ pages or block them from their newsfeed” (171). Participants developed and followed their own norms for understandability and moral evaluation but did not agree with the totality of their online social network on the validity of such norms. Similarly, Rost, Stahel, and Frey (2016) replicated an SNS-like environment to observe whether user anonymity favored aggression and outrage when discussing controversial issues. Their findings suggest not only a lack of relation between anonymity and aggression; (moral) outrage appeared to be a form of norm enforcement by the participants, who felt entitled to punish standpoints they deemed inappropriate or wrong – an entitlement that goes together with users’ identifiability. In this sense, their findings see outrage as a form of *sousveillance* and, while the result of frustration, not necessarily an irrational behavior. If anything, the apparent overstatement of one’s own values and standards through outrage and aggression is a result of human sensitivity to normatively relevant issues coupled with a lack of *shared* values and standards.

To conclude, context collapse and the SNSs’ interfaces, which exploit users’ normative sensitivity to their values, beliefs, and standards for right and wrong behavior, are key contributors to frustration, polarized aggression, and outrage. This effect of SNSs’ deliberate interface structure distorts our natural reliance on the context-dependent norms and values that ground our understanding and evaluation of other people – norms that are absent due to context collapse, which embodies one of the most radical shifts from face-to-face and social media-based interaction. Not just the moral evaluation of other people’s actions (our reactive attitudes) but the mere understanding of what a piece of behavior or utterance means is decontextualized, and therefore groundless. The transformation of social cognition and interaction and the rise of aggression and incivility on SNSs like Facebook and Twitter is explainable because of frustration, a systematized tendency to misunderstand others’ behavior due to a lack of a shared normative frame.

5. How Context Collapse Nudges Epistemic Injustice

So far, my argument possesses the potential to explain the rise in hostility in online interaction by understanding a key difference between social cognition in online

and offline environments. However, it is worth noting that this SNS-based dynamic enables a specific kind of injustice. I will now argue that, by the same logic by which SNSs such as Facebook and Twitter frustrate social cognition due to the lack of a shared frame, they also nudge epistemic, and specifically testimonial, injustice. In other words, the deliberate design choices of these platforms not only cause frustration in the understanding of others but can also be considered a key element in the mistreatment of others as competent knowers, speakers, and agents online.

Fricker (2007, 2016) defines testimonial injustice as a credibility deficit that a speaker suffers from due to the bias of the hearer; it occurs “when the level of credibility attributed to a speaker’s word is reduced by prejudice operative in the hearer’s judgement” (2016, 161). She proposes, as an example of testimonial injustice, the jury’s bias toward Tom Robinson, an African American, in Harper Lee’s novel *To Kill a Mockingbird*. The lack of trust experienced by the (all-white) jury of the trial toward the defendant is represented as a lack of credibility when it comes to the defendant’s account of what happened and his feelings toward the victim of the crime he is being blamed for. This lack of credibility depends not on the credibility of the defendant’s speech acts in themselves but on the jury’s prejudice toward the defendant as an African American, which has nothing to do with the sincerity, veracity, or credibility of the speech act as such. Testimonial injustice occurs when, for reasons unrelated to the speech act in question, a speaker’s credibility is downplayed or denied.

This is exactly the kind of position that SNSs like Facebook and Twitter, through the decontextualization of their users and their interactions, put people in – with one important exception. If typical cases of testimonial injustice occur due to a systematized belief or (dis)value (typically a bias toward a minoritarian group) shared within a community or social structure, in the case of testimonial injustice on SNSs, a systemic credibility deficit toward a speaker due to the hearer’s prejudice occurs because of a lack of any shared belief, value, or standards among the community of social media users. Rather, the prejudice stands because, due to context collapse, and with the encouragement of the platform’s interface design, the hearer needs to rely on their own beliefs, values, and (perceived) standards to judge others as competent or credible. If our mutual understandability, social evaluability, and capacity for interaction are, under ordinary circumstances, framed by socially shared values and norms, and if on SNSs like Facebook and Twitter, due to context collapse and their user-focused design, interaction is frustrated and groundless, then users’ evaluations of others as competent agents, knowers, and speakers are systematically prejudiced.¹⁵ Because understandability, accountability, *and* credibility standards are normally grounded in the shared practices of a community, and because many social media platforms do not provide any of such standards and actively encourage

¹⁵ In a typical case of testimonial injustice, the person committing the injustice is very often in a relation of power toward the victim – which is one of the reasons why it typically occurs, especially in its more systemic instances, toward marginalized groups. However, this criticism can be answered by underlining that SNSs, through extreme user focus, favor the impression of entitlement and empowerment in their users, which can come at the cost of acknowledging others’ competence as speakers and knowers. Epistemic injustice online is particularly favored by the platforms’ design, and thus it has a higher likelihood to occur, especially toward marginalized groups (Phillips and Milner 2021) due to the designers’ deliberate design choices.

their users to interpret the world through their own values and standards as part of their business model, SNSs end up nudging testimonial injustice, for they systematically favor a credibility deficit through context collapse and by encouraging individual users to interpret others' speech acts and actions through their own values and standards rather than through shared ones. In a sense, this is the result of SNSs' focusing their designs on decontextualized individuals, rather than seeing users as members of a community.

As Gunn (2020) argues, responsible agency meant to promote productive (or, in our case, not frustrated) communication and mutual understanding needs to be sensible to the context that mediates our conception of one another as competent agents and communicators. In this sense, the normative (both epistemic and moral) standards for individual behavioral and belief formation regulation need to be grounded in a joint commitment of all agents involved in the interaction to mutually acknowledged ends. The existence of shared epistemic and moral norms and values is a key enabler of functional communication and understanding; and if contravening such grounding norms can lead to injustices in the treatment of others as (in)competent agents and knowers, their absence impedes the just and appropriate treatment of others because there are no clear and shared standards for assessing what any piece of behavior means, let alone its appropriateness, veracity, or sincerity. If a speaker's and knower's understandability, accountability, and credibility also importantly depend on practice-specific standards, and if those standards are not just missing but the individual user is nudged to assume their own (and possibly inappropriate ones), then the likelihood of unjust discrediting of others as competent knowers, speakers, and agents becomes not just high but systematic. And, most importantly, this is a direct result of context collapse and the extreme user-centeredness of these SNSs' interfaces, as deliberate design choices.

6. Concluding Thoughts

In this paper, I argued that the lack of a normative infrastructure capable of framing mental state attribution in some SNSs has significantly negative consequences for online interaction. When there are no norms for social interaction online due to context collapse, our competence as social agents is hindered; we have no normative frame through which to understand others' beliefs, intentions, and desires. The lack of such a frame, which is often a deliberate choice by SNS designers, not only can lead easily to frustrated interaction and hostility online but can also be seen as a way of nudging one's own thoughts and values over other people's, leading to a systematic credibility deficit toward people who share different or contrasting values – in other words, nudging a systematic epistemic injustice.

My argument highlights the distortion of social cognition in online environments, while explaining how such transformation can lead to frustration, aggression, and epistemic injustice online. While an in-depth exploration of possible strategies to aptly tackle the negative impact of online context collapse on social media-based interaction would require a lengthy discussion, it is already possible to sketch two

parallel and complementary pathways, i.e., creating appropriate media literacy policies and implementing design changes for social media interfaces. Both these pathways ought to tackle the lack of shared norms within these online spaces: the goal, roughly sketched, would be to establish some mutually shared norms, or a *pragmatics*, for social media users that would facilitate mutual understanding and transform the social landscape of these platforms from a set of individual users to a community of people with shared values. Given that the user-centeredness of these platforms is largely responsible for context collapse and, in turn, systematic epistemic injustice, a media literacy program focused on understanding the presence and reasons of other people online – see Phillips and Milner (2021), e.g. – would be helpful in mitigating the negative effect of context collapse. On the other hand, it might be possible to make some design implementations that would counterbalance the tendentially egocentric configuration of these platforms. An example would be the implementation of pop-up notifications that might prompt users to respond reflectively, rather than instinctively, to posts they may want to verbally sanction – a sort of “think twice” function that would help users keep in mind the diverse backgrounds of the people they may interact with, and hopefully mitigate the effects of context collapse. The overall goal, for both an adequate media literacy policy and a set of design interventions, is to favor the establishment of shared norms and values across all users, whose awareness and respect can constitute a frame through which to promote mutual understanding and interaction.

References

- Andrews, Kristin. *Do Apes Read Minds?: Toward a New Folk Psychology*. Cambridge, MA: MIT Press, 2012.
<https://doi.org/10.7551/mitpress/9283.001.0001>
- Andrews, Kristin. “The Folk Psychological Spiral: Explanation, Regulation, and Language.” *Southern Journal of Philosophy* 53, no. S1 (2015): 50–67.
<https://doi.org/10.1111/sjp.12121>
- Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. “Assessing the Russian Internet Research Agency’s Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017.” *Proceedings of the National Academy of Sciences* 117, no. 1 (January 7, 2020): 243.
<https://doi.org/10.1073/pnas.1906420116>
- Beam, Michael A, Jeffrey T. Child, Myiah J. Hutchens, and Jay D. Hmielowski. “Context Collapse and Privacy Management: Diversity in Facebook Friends Increases Online News Reading and Sharing.” *New Media & Society* 20, no. 7 (July 2018): 2296–2314.
<https://doi.org/10.1177/1461444817714790>
- Bratman, Michael E. *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press, 2014.

-
- Büchi, Moritz, Eduard Fosch Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, Shruthi Velidi, and Salome Viljoen. "Chilling Effects of Profiling Activities: Mapping the Issues." *SSRN Electronic Journal* (2019).
<https://doi.org/10.2139/ssrn.3379275>
- Burr, Christopher, Nello Cristianini, and James Ladyman. "An Analysis of the Interaction between Intelligent Software Agents and Human Users." *Minds and Machines: Journal for Artificial Intelligence, Philosophy and Cognitive Science* 28, no. 4 (2018): 735–74.
<https://doi.org/10.1007/s11023-018-9479-0>
- Costa, Elisabetta. "Affordances-in-Practice: An Ethnographic Critique of Social Media Logic and Context Collapse." *New Media and Society* 20, no. 10 (2018): 3641–56.
- De Jaegher, Hanne, and Ezequiel Di Paolo. "Participatory Sense-Making: An Enactive Approach to Social Cognition." *Phenomenology and the Cognitive Sciences* 6, no. 4 (2007): 485–507.
- Eyal, Nir. *Hooked: How to Build Habit-Forming Products*. New York: Portfolio Penguin, 2014.
<https://books.google.nl/books?id=YNZZoAEACAAJ>
- Fernandez Castro, Victor. "Regulation, Normativity and Folk Psychology." *Topoi* 39, no. 1 (2020): 57–67.
<https://doi.org/10.1007/s11245-017-9511-7>
- Figà Talamanca, Giacomo. "Joint Action without Mutual Beliefs." *KRITERION – Journal of Philosophy* 35, no. 1 (March 30, 2021): 47–70.
<https://doi.org/10.1515/krt-2021-0004>
- Flaxman, Seth, Sharad Goel, and Justin M. Rao. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80, no. S1 (January 1, 2016): 298–320.
<https://doi.org/10.1093/poq/nfw006>
- Fox, Jesse, and Jennifer J. Moreland. "The Dark Side of Social Networking Sites: An Exploration of the Relational and Psychological Stressors Associated with Facebook Use and Affordances." *Computers in Human Behavior* 45 (April 2015): 168–76.
<https://doi.org/10.1016/j.chb.2014.11.083>
- Fricker, Miranda. "Epistemic Injustice and the Preservation of Ignorance." In *The Epistemic Dimensions of Ignorance*, edited by Martijn Blaauw and Rik Peels, 160–77. Cambridge: Cambridge University Press, 2016.
<https://doi.org/10.1017/9780511820076.010>
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007.
- Gallagher, Shaun. "The Practice of Mind: Theory, Simulation or Primary Interaction?" *Journal of Consciousness Studies* 8, no. 5–7 (2001): 83–108.
- Gallagher, Shaun. "Understanding Interpersonal Problems in Autism." *Philosophy, Psychiatry, and Psychology* 11, no. 3 (2004): 199–217.
<https://doi.org/10.1353/ppp.2004.0063>
- Gallese, Vittorio. "Mirror Neurons, Embodied Simulation, and the Neural Basis of Social Identification." *Psychoanalytic Dialogues* 19, no. 5 (October 13, 2009): 519–36.
<https://doi.org/10.1080/10481880903231910>
- Gilbert, Margaret. *Joint Commitment: How We Make the Social World*. New York: Oxford University Press, 2013.
- Goldman, Alvin I. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press, 2006.

- Gunn, Hanna Kiri. "How Should We Build Epistemic Community?" *Journal of Speculative Philosophy* 34, no. 4 (2020): 561–81.
<https://doi.org/10.5325/jspecphil.34.4.0561>
- Hagey, Keach, and Jeff Horwitz. "Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead". September 15, 2021.
<https://www.wsj.com/articles/facebook-algorithm-change-zucker-berg-11631654215>
 Accessed December 15, 2021.
- Harel, Tal Orian, Jessica Katz Jameson, and Ifat Maoz. "The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict." *Social Media + Society* 6, no. 2 (April 2020).
<https://doi.org/10.1177/2056305120913983>
- Hildebrandt, Mireille. "Defining Profiling: A New Type of Knowledge?" In *Profiling the European Citizen*, edited by Mireille Hildebrandt and Serge Gutwirth, 17–45. Dordrecht: Springer Netherlands, 2008.
https://doi.org/10.1007/978-1-4020-6914-7_2
- Hildebrandt, Mireille, and Serge Gutwirth. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. [New York]: Springer, 2008.
http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=2877919&custom_att_2=simple_viewer
- Horne, Christine. "The Enforcement of Norms: Group Cohesion and Meta-norms." *Social Psychology Quarterly* 64, no. 3 (2001): 253–66.
<https://doi.org/10.2307/3090115>
- Hutto, Daniel D. "Folk Psychology as Narrative Practice." *Journal of Consciousness Studies* 16, no. 6–8 (2009): 6–8.
- Kiesler, Sara, Jane Siegel, and Timothy W. McGuire. "Social psychological aspects of computer-mediated communication." *American psychologist* 39, no.10 (1984): 1123–1134.
<https://doi.org/10.1037/0003-066X.39.10.1123>
- Kim, Eun-Mee, and Jennifer Ihm. "Online News Sharing in the Face of Mixed Audiences: Context Collapse, Homophily, and Types of Social Media." *Journal of Broadcasting & Electronic Media* 64, no. 5 (December 1, 2020): 756–76.
<https://doi.org/10.1080/08838151.2020.1835429>
- Kirkpatrick, David. *The Facebook Effect: The inside Story of the Company That Is Connecting the World*. New York: Simon & Schuster, 2010.
<http://catdir.loc.gov/catdir/enhancements/fy1011/2009051983-s.html>
- Kowalski, Robin M., Gary W. Giumetti, Amber N. Schroeder, and Micah R. Lattanner. "Bullying in the Digital Age: A Critical Review and Meta-analysis of Cyberbullying Research among Youth." *Psychological Bulletin* 140, no. 4 (July 2014): 1073–1137.
<https://doi.org/10.1037/a0035618>
- Langos, Colette. "Cyberbullying: The Challenge to Define." *Cyberpsychology, Behavior, and Social Networking* 15, no. 6 (June 2012): 285–89.
<https://doi.org/10.1089/cyber.2011.0588>
- Mameli, Matteo. "Mindreading, Mindshaping, and Evolution." *Biology & Philosophy* 16, no. 5 (November 2001): 595–626.
<https://doi.org/10.1023/A:1012203830990>

-
- Marin, Lavinia. "Sharing (Mis) Information on Social Networking Sites. An Exploration of the Norms for Distributing Content Authored by Others." *Ethics and Information Technology* 23, no. 3 (September 2021): 363–72.
<https://doi.org/10.1007/s10676-021-09578-y>
- Marwick, Alice E., and danah boyd. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13, no. 1 (February 2011): 114–33.
<https://doi.org/10.1177/1461444810365313>
- Marwick, Alice E., and danah boyd. "Networked Privacy: How Teenagers Negotiate Context in Social Media." *New Media & Society* 16, no. 7 (November 1, 2014): 1051–67.
<https://doi.org/10.1177/1461444814543995>
- Masullo Chen, Gina, Ashley Muddiman, Tamar Wilner, Eli Pariser, and Natalie Jomini Stroud. "We Should Not Get Rid of Incivility Online." *Social Media + Society* 5, no. 3 (July 2019).
<https://doi.org/10.1177/2056305119862641>
- McGeer, Victoria. "Enculturating Folk Psychologists." *Synthese* 199, no. 1–2 (December 2021): 1039–63.
<https://doi.org/10.1007/s11229-020-02760-7>
- McGeer, Victoria. "Mind-Making Practices: The Social Infrastructure of Self-Knowing Agency and Responsibility." *Philosophical Explorations* 18, no. 2 (May 4, 2015): 259–81.
<https://doi.org/10.1080/13869795.2015.1032331>
- McGeer, Victoria. "Scaffolding Agency: A Proleptic Account of the Reactive Attitudes." *European Journal of Philosophy* 27, no. 2 (2019): 301–23.
- Moor, Peter J., Ard Heuvelman, and Ria Verleur. "Flaming on YouTube." *Computers in Human Behavior* 26, no. 6 (November 2010): 1536–46.
<https://doi.org/10.1016/j.chb.2010.05.023>
- Nichols, Shaun, and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. New York: Clarendon Press, 2003.
<https://doi.org/10.1093/0198236107.001.0001>
- Osler, Lucy. "Taking Empathy Online." *Inquiry: An Interdisciplinary Journal of Philosophy* (March 15, 2021): 1–28.
<https://doi.org/10.1080/0020174X.2021.1899045>
- Pacilli, Maria Giuseppina, Michele Roccato, Stefano Pagliaro, and Silvia Russo. "From Political Opponents to Enemies? The Role of Perceived Moral Distance in the Animalistic Dehumanization of the Political Outgroup." *Group Processes & Intergroup Relations* 19, no. 3 (May 2016): 360–73.
<https://doi.org/10.1177/1368430215590490>
- Peters, Uwe. "The Complementarity of Mindshaping and Mindreading." *Phenomenology and the Cognitive Sciences* 18, no. 3 (July 2019): 533–49.
<https://doi.org/10.1007/s11097-018-9584-9>
- Phillips, Whitney. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press, 2015.
- Phillips, Whitney, and Ryan M. Milner. *You Are Here: A Field Guide for Navigating Polarized Speech Conspiracy Theories, and Our Polluted Media Landscape*. Cambridge, MA: MIT Press, 2021.

- Rains, Stephen A., Kate Kenski, Kevin Coe, and Jake Harwood. "Incivility and Political Identity on the Internet: Intergroup Factors as Predictors of Incivility in Discussions of News Online: Incivility and Political Identity Online." *Journal of Computer-Mediated Communication* 22, no. 4 (July 2017): 163–78.
<https://doi.org/10.1111/jcc4.12191>
- Rini, Regina. "Fake News and Partisan Epistemology." *Kennedy Institute of Ethics Journal* 27, no. 2S (2017): E-43–E-64.
<https://doi.org/10.1353/ken.2017.0025>
- Rost, Katja, Lea Stahel, and Bruno S. Frey. "Digital Social Norm Enforcement: Online Firestorms in Social Media." *PLOS ONE* 11, no. 6 (June 17, 2016): e0155923.
<https://doi.org/10.1371/journal.pone.0155923>
- Selinger, Evan, and Woodrow Hartzog. "Facebook's Emotional Contagion Study and the Ethical Problem of Co-opted Identity in Mediated Environments Where Users Lack Control." *Research Ethics* 12, no. 1 (January 2016): 35–43.
<https://doi.org/10.1177/1747016115579531>
- Sellars, Wilfrid S. "Empiricism and the Philosophy of Mind." *Minnesota Studies in the Philosophy of Science* 1 (1956): 253–329.
- Strawson, Peter Frederick. *Freedom and Resentment and Other Essays* (1st ed.). Routledge, 2008.
<https://doi.org/10.4324/9780203882566>
- Torrance, Steve, and Tom Froese. "An Inter-enactive Approach to Agency: Participatory Sense-Making, Dynamics, and Sociality." *Humana.Mente* 15 (2011): 21–53.
- Vincent, Sarah, Rebecca Ring, and Kristin Andrews. "Normative Practices of Other Animals." In *The Routledge Handbook of Moral Epistemology*, edited by Aaron Zimmerman, Karen Jones, and Mark Timmons, 57–83. Abingdon: Routledge, 2018.
- Vitak, Jessica. "The Impact of Context Collapse and Privacy on Social Network Site Disclosures." *Journal of Broadcasting & Electronic Media* 56, no. 4 (October 2012): 451–70.
<https://doi.org/10.1080/08838151.2012.732140>
- Vitak, Jessica, Stacy Blasiola, Eden Litt, and Sameer Patil. "Balancing Audience and Privacy Tensions on Social Network Sites: Strategies of Highly Engaged Users." *International Journal of Communication* 9 (May 1, 2015).
<https://ijoc.org/index.php/ijoc/article/view/3208>
- Wang, Yiran, and Gloria Mark. "Engaging with Political and Social Issues on Facebook in College Life." In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 433–45. Portland, OR, USA: ACM, 2017.
<https://doi.org/10.1145/2998181.2998295>
- Westra, Evan. "Folk Personality Psychology: Mindreading and Mindshaping in Trait Attribution." *Synthese* 198, no. 9 (2020): 8213–32.
- Zahavi, Dan. "Beyond Empathy: Phenomenological Approaches to Intersubjectivity." *Journal of Consciousness Studies* 8, no. 5–7 (2001): 151–67.
- Zahavi, Dan. "Empathy and Direct Social Perception: A Phenomenological Proposal." *Review of Philosophy and Psychology* 2, no. 3 (2011): 541–58.
<https://doi.org/10.1007/s13164-011-0070-3>
- Zawidzki, Tadeusz. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press, 2013.

Trust is in the (linguistic) details: policing digilect as a communication practice

Applying Instagram in two Hungarian law enforcement units as a possible communication channel

The use of social network sites by law enforcement agencies takes place between the police as a service provider and the members of the public. Site administrators need to find a communication style that is understandable to the broadest audience. In the course of my research, I examined the entries created on the Instagram profiles of the Budapest Metropolitan Police Headquarters (Budapesti Rendőrfőkapitányság, hereinafter: BRFK) and the Hungarian Police (Magyar Rendőrség, hereinafter: MR) in two sampling periods (sampling took place in the first month of both profiles, and a year later) using the methodology of content and discourse analysis. Comparison of the results has shown that there is a more casual style of gaining the trust of the population. Appropriate and regular partnership communication with civilians can be used effectively in crime prevention and detection, and can also have positive effects on security.

Keywords: *Hungarian Police, police communications, organisational strategy, social media, Instagram*

Author Information

Erna Uricska, Corvinus University of Budapest
<https://orcid.org/0000-0002-5371-0650>

How to cite this article:

Uricska, Erna . “Trust is in the (linguistic) details:
policing digilect as a communication practice”.
Információs Társadalom XXIII, no. 2 (2023): 146–162.
== <https://dx.doi.org/10.22503/inftars.XXIII.2023.2.9> ==

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

Traditional communication channels have been applied by organisations to inform citizens for decades. Until the 2000s, members of society were able to access police communications through traditional media (e.g., television, radio and newspapers), but since the 2010s, various community platforms, such as profiles operated by law enforcement agencies (see law enforcement community sites), have become widespread worldwide, including in Hungary. In recent years, social media has emerged as an important and modern communication tool and applied by individuals, groups and organisations such as law enforcement units.

Organisations are established to plan, implement or perform specific tasks. Communication is a basic requirement for the proper functioning of organisations, and the concept of organisational communication can be approached from two directions: internal organisational communication and external organisational communication (Molnár 2018; Winarso 2018).

Traditional communication channels have been applied by organisations to broadcast essential information to citizens. Nowadays, social media has a pivotal role as individuals, groups, entrepreneurs, conventional media and private institutions have adopted it to gather and share information with their audiences (Carr and Hayes 2015; Heravi and Harrower 2016). It also plays a crucial role in the communication practices of governmental agencies and organisations that interact directly with individuals (Herfort et al. 2014).

With social media's widespread use and growing popularity, interest in its potential has also increased in different areas of life, such as marketing (Veszelszki 2019), organisational communication (Tajudeen et al. 2017; Bullock 2018) and education (Putra 2021), with innovations in techniques being introduced, namely content marketing and organic marketing (Veszelszki 2019). One of the pivotal advantages of social media is that it supports open and two-way communication, enabling different organisations to engage with their consumers and assess their needs, thus adapting the content of the site and the services offered to them (Tajudeen et al. 2017). Therefore, the aim of the online process is to keep in touch with customers and interested parties, and to develop a proper communication strategy (Stenger 2014). It is obvious that an organisational strategy is required to define long-term objectives, action lines and tasks, and to be effective in achieving organisational goals (Uricska 2021).

2. Law enforcement and communication

The required tasks of police personnel are declared in the Fundamental Law of Hungary. According to Article 46, section (1): *'The fundamental duty of the Police shall be the prevention and investigation of criminal activities, as well as the protection of public security, public order and of the state borders. The Police shall participate in efforts to prevent illegal immigration.'* As Géza Finszter (2018) highlights, *'in Hungary the Fundamental Law separates two basic tasks of the police: on the one hand, the prevention and detection of crime, and on the other, the protection of public safety and*

public order. The duality of law enforcement organisations is also based on these two-fold tasks: as an administrative authority, police protect public order and public safety, and on the other hand, as a criminal investigation service, they prepare cases related to criminal justice proceedings' (Finszter 2018, 113). As a consequence of these required tasks, the communication of the police is primarily social communication, since their tasks include crime prevention, crime detection, preparation of justice, maintenance and restoration of public order and security, and accident prevention (Molnár 2018).

In modern democracies, the relationship between the citizens and the police is influenced and affected by the quality of their communication to a great extent. In the most general terms, the police provide services to civilians, and these services include a wide range of communication activities.

A key aspect of organisational communication is that it does not have to be a chain of random events, but it has to be *'the purposeful use of communication by the organization to fulfill its mission'* (Hallahan et al. 2007, 3). For that reason, law enforcement communication should rather be a consciously planned process, and this process should be explained in detail. In this regard, *'a law enforcement communication strategy is a short-, medium-, or long-term and conscious design of how law enforcement agencies should present themselves to the outside world, and facilitate the internal flow of information, and continuous, high-quality and efficient operation'* (Molnár 2018, 9).

The main deficiency of a centralised police force is that it perceives the needs of the society in relation to public safety inaccurately and distantly, and responds to them slowly (Christián et al. 2021). However, since 2010, significant changes have taken place in the professional and structural system of the Hungarian Police and the Communication Service has been set up (Uricska 2020). Despite the strict hierarchical structure of the police, almost independent positions appeared in the field of communication and were directly channelled to the high commissioner of the MR.

3. Law enforcement, communication and social media

With the emergence of social networking sites, other types of crime and communication objective have emerged in the work of law enforcement, and the methods that police use to communicate with the public have changed, too (Lee and McGovern 2013; Fielding 2021). On online community interfaces, there is a chance for bilateral and informal communication; therefore the police need to find a style that is understandable to the broadest audience possible (Uricska 2021) if they want to reach every member of the community with their messages.

Organisations apply social media in their everyday work; it is one of the most powerful community engagement tools they possess in the 21st century. Informing, educating and engaging the members of communities have never been as simple as it is today. Moreover, developing an organisational communications strategy can be one of the most important ways to build a stronger brand.

Instagram was launched in 2010 by an independent private company; it began as an app only available on iPhones (Highfield and Leaver 2016). The application had a very small userbase, and non-mobile users could not use it. Now it is owned by Meta, but in the beginning, Instagram was run by an independent private company (Highfield and Leaver 2016). Instagram, as its name suggests – the word is a portmanteau of *instant* and *camera* – contains more visual communication elements than Twitter or Facebook, and it is impossible to create a post on Instagram without a picture or video, while a post can be created on Facebook with or without visual content.

There is tremendous potential for Hungarian law enforcement units to use these sites (for example, for crime prevention purposes). For instance, statistics (URL 1) show that there are 2.6 million Instagram users in Hungary, of which 860,000 fall between the ages of 18 and 24, although this age group includes only 7 years of age. They are followed in group size by the 25–34-year-old users, of which there are 770,000 users. Meanwhile, the over 55s have only 139,000 registered users on Instagram (Uricska 2023).

In Hungary, the rise of social media coincided with the expansion in police organisational communications around the 2010s. Before the appearance of social media, the Hungarian police organisations informed the general public mainly through the traditional channels of media including police.hu, the official webpage of the MR. Taking advantage of the opportunities offered by social networking sites, the Hungarian law enforcement organisations have also now appeared on social media platforms, representing their organisations and seeking more direct contact with members of society.

The MR has had an official website since 2000 (police.hu), a Twitter account since 2016, a police_hu page on Instagram since 2019 (Uricska 2021), and a Facebook page since 2020. The MR can be regarded as a pioneer within the organisations of public administrations, as the Government of Hungary has been present on Facebook since 2011, and has also had an Instagram profile since 2020.

In this study, I observed the Instagram profiles of two Hungarian law enforcement units, to see if Instagram is a possible communication channel for contemporary police communication strategy, regarding short-term goals and the achieved results. The first organisation was the abovementioned MR; the other was the BRFK. The BRFK is the administrative organisation of the MR. Its tasks relate to the administrative and operational tasks of police activities and police personnel within the territory of Budapest.

The main focus was on the verbal and visual content on Instagram of these two Hungarian police units and it was observed whether an organisational strategy could be detected or witnessed in police organisational communications; more precisely, I explored the ways in which police communications enhanced dialogue with the members of the community.

The first sampling period of the Instagram profiles of the BRFK (URL 2) and the MR (URL 3) was the first month of the profiles; this was between 6 December 2017 and 6 January 2018 in the case of the BRFK, and between 3 July 2019 and 3 August 2019 in the case of the MR. The sampling period was repeated a year later for a month again, including both profiles.

4. Research questions and hypotheses

The purpose of the research is to explore whether a coherent organisational strategy and communication practice can be identified for the Hungarian police organisations by observing two profiles through their contents and activities on Instagram.

The MR first appeared on Instagram on 3 July 2019 with the profile of *police_hu*; the other Hungarian law enforcement agency, the BRFK, appeared on the platform (*brfk_budapest_police_hu*) on 6 December 2017. In order to complete the research, the existence and the quality of the digital police communication strategy were examined and the following research questions were formulated:

RQ1) What kind of organisational communication strategy can be detected in the content, themes and posting frequency of the Instagram profiles of the two law enforcement units?

RQ2) Can the same organisational communication strategy be detected on the profiles of the two law enforcement units?

RQ3) If so, in what form (visual content, verbal content, hashtags)?

During my research I tried to test two hypotheses based on the theoretical background. Therefore, my hypotheses for the research questions are as follows:

- 1) To reach the members of the community, law enforcement units appeared deliberately on Instagram in Hungary; however, no coherent organisational communication strategy can be detected in the content, themes and posting frequency of the Instagram profiles of the two law enforcement units.
- 2) There is a more informal language variant on Instagram, the *policing digilect*, a communication practice that is closer to verbalism in its style features and has special linguistic traits.

5. Research design and methodology

Content analysis and Instagrammatics methods were applied to complete the study (Highfield and Leaver 2016). The starting point in studying the communicational strategy on Instagram was built on the established research method of content analysis. According to Berelson, content analysis is '*a research technique for the objective, systematic and quantitative description of the manifest content of communication*' (Berelson 1952, 18). Babby (2001) defined content analysis as '*simply the study of recorded human communication*' (Babby 2001, 352) and stated that there are two possible ways of analysing content: '*when encoding the (1) manifest (surface) content or (2) the latent content of a message*' (Babby 2001, 352). Additionally, Instagrammatics were applied to examine this primarily visual platform '*to track and study Instagram activity, content, and practices*' (Highfield and Leaver 2016, 5). The contents of

the platforms also offered an opportunity to observe an in-depth understanding of community engagement and communication (Varis 2016).

To complete the study, two sampling periods were observed: the first month of the appearance of the Instagram profiles of the BRFK and the MR in the first phase, and a repeat of the observations exactly a year later (N = 127 entries) in the second phase.

Units of analysis	Budapest Metropolitan Police Headquarters (brfk_budapest_police_hu)	Hungarian Police (police_hu)
Total posts observed (N=127)	5	122

Table 1. Sample selection from the Instagram profiles of Hungarian law enforcement organisations (own edition)

During these periods, in order to find answers to RQ1 and RQ2, the following categories were examined:

- 1) the first day of the appearance (time interval up to now)
- 2) the number of followers
- 3) the theme and mood of the first post on the profiles
- 4) the themes of the first sampling period
- 5) the themes of the second sampling period.

In this study, the manifest content is observed and the results are presented in numerical style (Berelson 1952; Krippendorff 2004).

6. Results

6.1. The first day of the appearance and the number of followers

The MR first appeared on 3 July 2019 on Instagram with its profile (police_hu), and the account had 74,600 followers on the date of the observation (9 October 2021). The other Hungarian law enforcement agency was the BRFK, which appeared on the platform (brfk_budapest_police_hu) on 6 December 2017; the number of followers was 3,528 on 9 October 2021 (Table 2).

Units of analysis	Budapest Metropolitan Police Headquarters (brfk_budapest_police_hu)	Hungarian Police (police_hu)
Number of followers on the date of sampling (9 October 2021)	3,528	74,600
Number of posts on the date of sampling (9 October 2021)	392	1,232

Table 2. The number of followers of the two profiles (own edition)

Despite the fact that the MR appeared one and a half years later, the organisation gained twenty times more Instagram followers than the BRFK. Since its appearance on Instagram, the MR had created 1,232 posts, while the other organisation had created only 392. Regarding the frequency of posting, the MR applied a more conscious communication practice as part of the communication strategy.

6.2. The theme and mood of the first post on the profiles and the first sampling period (content and themes)

Units of analysis	Budapest Metropolitan Police Headquarters (brfk_budapest_police_hu)	Hungarian Police (police_hu)
1st sampling period	6 Dec 2017–6 Jan 2018	3 Jul 2019–3 Aug 2019
Total number of posts (n1)	3	82
Promoting police profession	1	30
Display of police vehicles (water police, motor police, mounted police)		19
#thenandnow (hashtag)		6
Crime prevention		7
#proudofoyou (hashtag)		6
Promoting the emergency number 112		7
Holidays	2	
Others		7

Table 3. Sample selection from the Instagram profiles of Hungarian law enforcement organisations (own edition)

The first sampling period was the first month (6 December 2017–6 January 2018) when the BRFK appeared on Instagram, and during this time the organisation created only three entries. The first post was uploaded on 6 December and consists of a picture of Santa Claus on a bicycle as the post was created on Santa Claus Day. There was neither a greeting to the followers on the profile nor other organisational comments before or after it (Figure 1). Only two short sentences were written next to the post by the administrator (including a grammatical mistake in Hungarian): *Van aki ilyenkor is kerékpározik. Vigyázzunk rájuk közlekedés közben!* ('Some people ride bikes at this time of year. Let's watch out for them on the road!' [the author's translation]). In Hungarian, between the words *Van* and *aki* a comma is required. These sentences are followed by hashtags that appear to be random: #BRFKInformációsPortál ('the information portal of BRFK') #BRFK #adventitanácsok

In the first sampling period, the MR posted 82 entries and the topics (Table 3) were as follows:

- promoting the police profession and #proudfyou (hashtag, used in English)
- display of police vehicles (water police, motor police, mounted police)
- #thenandnow (hashtag, used in English)
- crime prevention
- promoting the emergency number 112.

The profile of the MR is characterised by predictable daily content and quality images (e.g., resolution and saturation). Three constant hashtags are used as the trademark of the site: #rendőrség #police #hivatás. The meaning of the first hashtag is ‘police’, and the meaning of the third one is ‘profession’ in Hungarian.

Though 82 entries can be seen as a high number within a month, the organisation was eager to communicate to the public with more than two posts an average per day. Due to an interview that was implemented by one of the administrators of police_hu on 23 April 2021, the profile reached 10,000 followers in a week.

6.3. The themes of the second sampling period

Units of analysis	Budapest Metropolitan Police Headquarters (brfk_budapest_police_hu)	Hungarian Police (police_hu)
2nd sampling period	6 Dec 2018–6 Jan 2019	3 Jul 2020–3 Aug 2020
Total number of posts (n2)	2	40
Promoting police profession	1	10
Display of police vehicles (water police, motor police, mounted police)		17
#thenandnow (hashtag)		4
Responsible pet ownership		6
Promoting the emergency number		1
Holidays	1 (+1)	
Others		2

Table 4. Sample selection from the Instagram profiles of the Hungarian law enforcement organisations (own edition)

In the second sampling period (Table 4), a similar content was created by the BRFK as in the first one; only two entries were posted without observable policing content. Through '*objective, systematic and quantitative description of the manifest content of communication*' (Berelson 1952, 18), it can be stated in the case of the BRFK that no organisational strategy can be observed by analysing the content on Instagram. If the organisation has such a strategy, the administrators do not apply it with predictable and daily content, or coordinated verbal and visual elements on the site. However, the annual budget for content management was not known in the case of the police organisations.

In the case of the MR, 40 entries were created. The themes and projects remained almost the same, except that the theme of *crime prevention* was replaced by the theme of *responsible pet ownership*. The themes of *promoting the police profession* and *#proudofoyou* hashtags and *the display of police vehicles* were the central themes in terms of visual content; the site might have an extremely important role in the recruitment process and making the police career seem more attractive to citizens. As proven by data (Tables 3 and 4), the organisation has a kind of strategy including special themes and projects, and there is a sense of organisational mission behind the content. Since the sampling periods, the Instagram profile of the MR has been awarded two silver grades at *Creative Prisma* (URL 6), so the mission, the textual content and the images of the profile are appreciated and well-known by communication experts. The title of the project was '*You have shown the human face of the police*'.

The visual and verbal content of *crime prevention* posts and the conscious application of the *emergency number 112* posts can have far-reaching effects in saving lives and helping those who are in a state of emergency. The educational messages of the posts are not served directly, but they are hidden either in shocking content or in funny texts, and they likely have much greater impact among the members of Generation Z as authoritarian education no longer works in the 21st century (Kulcsár 2020; Valintine 2019).

6.4. Use of language: policing digilect as a communication practice

As the results presented in relation to RQ1 and RQ2 show, the BRFK (*brfk_budapest_police_hu*) created a total of five entries in the two sampling periods, of which two entries promoted the police profession and three entries contained the topic of holidays: Santa Claus Christmas and Christmas greetings from the organisation. In parallel with this, the MR (*police_hu*) introduced the work of the police with Kristóf Gál reciting Belga's poem 'Police Work', and created 122 entries in the sampling periods. They apply three constant hashtags: *#police #rendőrség #hivatus* and the administrators of the site implement *minimum response time* and *have a real dialogue between the police and the community members*.

In order to find answers to my third research question, as a linguistic methodological framework, netlinguistics (Posteguillo 2002) was analysed in relation to Veszelszki's term *digilect* (Veszelszki 2013, 2017a, 2017b) in the posts and comment

sections. In the case of the police units, the impact of infocommunication technology can be observed. The lexical features of digilect can be found in the Instagram profile of police_hu. By Veszelszki's observation (2013), the foreign (especially English) language effect is very strong in Hungarian-language internet communication. The above-mentioned features, due to their frequent occurrence – for example, foreign-language influence, breaking taboos, emojis, a neologism, a net-specific acronym – were detected in the sampling period (see Table 5), albeit slang words and expressions, as well as verbal aggression posted by the organisation could not have been observed on the policing profiles on Instagram.

Features	Examples
Foreign-language influence	<i>Brothers in arms. Drive safe. This is how we do it!, Thanks for the great performance to our colleagues from Heves County!</i> 🙌 🤝, <i>Sound on. Mornings like this.</i> 😊 <i>feeling, timelapse</i>
Neologisms	<i>COVID-19</i>
Net-specific acronyms	<i>RUTIN (Rendőrségi Útinformációs Rendszer, a police application to help people on the road)</i>
Letter substitution with number citizen's comment	<i>5let (< ötlet 'idea')</i>
Word multiplication	<i>stip-stop</i>
Word truncation	<i>heli ('helicopter')</i>
Breaking taboos	<i>My dad 🤔 ❤️. Te bekötöd!?</i> (‘Do you fasten [your seatbelt]?’), <i>Nos, kitaláljátok mi ez?</i> (‘Well, can you guess what it is?’), <i>buksi ('head', a synonym used for head, childish)</i>
Emojis	See the examples above.

Table 5. Features of digilect on the Instagram profile of Hungarian Police (own edition)

6.4.2. The special features of policing digilect

As on an international scale, numerous police organisations ‘have made a strategic decision to employ humour on social media to increase community engagement with their content’ (Wood and McGovern 2021, 305). This organisational goal can also be found in the forms of humour, puns and word-plays of the verbal content (see rows 1–2 in Table 6) on police_hu. In row 3 of Table 6, the word *iskolaőr* (‘school warden’) can be considered a *realia*. According to Klaudy (1997), *realia* is a generic term for objects that refer to a culture or language community and are unknown to other cultures.

Features	Examples
Humour	<p>VAKÁCIÓÓÓ! Gál Kristóf üzeni: figyeljünk a zebrákra (is)! Mert a nyári szünet alatt több gyerek van az utakon. 🕒📷👉🚗. ('HOOLIDAY! Kristóf Gál says: Watch out for zebras (too)! Because there are more children on the roads during the summer holidays.)</p> <p>Velünk senki sem cicózhat! 🐾🐱 ('Nobody can play with us!')</p>
Puns, word-plays	<p><i>mindenhol IS</i> ('everywhere too', a popular phrase in Hungarian in 2019–22)</p> <p><i>CEL-EB</i> (<i>eb</i> means a 'dog' in Hungarian, a dog is in the image, suggesting that the dog is a celebrity),</p> <p><i>Blue is the COLOUR</i>,</p> <p><i>Nézzetek csak, valaki merre SÜNdörög...</i> ('Watch out, someone is wandering...', <i>sün</i> is a 'hedgehog' in Hungarian, <i>sündörög</i> is a synonym for wandering)</p>
Realia	<i>iskolaőr</i> ('school warden')
Termini technici	<p><i>robbanószer</i> ('explosive material'), <i>112-es hívásfogadó operátor</i> ('112 emergency call operator'), <i>mentő</i> ('ambulance'), <i>ügyeletes</i> ('duty officer'), <i>vízi segélyhívó</i> ('water emergency'), <i>körzeti megbízott</i> ('police constable'), <i>lőmaradvány</i> ('remnants of ammunition')</p>
Poems, songs, film titles and computer games	<p><i>Belga: Rendőrmunka</i> ('Police Work' by Belga'), <i>Gyűrűk Ura</i> ('Lord of the Rings'), <i>Két torony</i> ('Two Towers'),</p> <p><i>A Király visszatér</i> ('The Return of the King')</p>
Metonymies	<p><i>Köszöntjük családunk új tagjait</i> 🕒📷 ('Welcome the new members of our family'), <i>kék család</i> ('blue family')</p>
Metaphors	<p><i>Ha Frodónak ilyen segítők lettek volna</i> ('If Frodo had had such helpers...', referring to police officers and the film <i>The Lord of the Rings</i>), <i>család</i> ('family')</p>

Table 6. Features of policing digilect on the Instagram profile of the Hungarian Police (own edition)

As the examples in Tables 4 and 5 show, *the posts and comments* (here mainly the posts) are “minimalist” in nature or they tend to be minimalist, by using abbreviations,

acronyms, emojis, and hashtags to help the reader quickly encode (create) and decode (absorb) text. As a result of the minimalist nature, writing time reduces and searching processes speed up' (Istók 2019, 89). It is explained in detail in Istók's concept of *minimalect* (Istók 2018).

As the observed profile is a law enforcement one, several termini technici were coded and centred on the topics of first aid, emergencies, guns and ammunition (see row 4 in Table 6). Metaphors and metonymies also help followers to engage. Metaphors importantly are not merely linguistic phenomena but a fundamental feature of the way the human cognitive system works (Lakoff 1995). In their work, Johansen et al. (2013) list three kinds of metaphor for trust: 'trust as a decision', 'trust as a performance' and 'trust as an uncontrollable force'. The word *család* ('family') belongs to the 'trust as a performance' metaphor group (see rows 5–6 in Table 6) as it refers to the smallest unit of society, and it suggests '*considerable control on behalf of the trustor in that a trustor is assumed to influence the trustworthiness of the trustee*' (Johansen et al. 2013, 7). The poems, songs, film titles and computer games also help followers get involved in the content, and get closer to the police by this police communication practice.

In conclusion, the results of the sampling periods show that *digilect* is a (digital) language variety that has specific quantitative and qualitative features. The concept of *policing digilect* can be introduced to name the language variety used in some specific online law enforcement community sites as a communication practice. By applying this new variety, proper language can be one of the tools with which police–community relations can be improved and rearranged in the long run.

As there is a growing need to establish and maintain the online relationship between the police and the members of the community, this platform can serve as a way to enhance the dialogue between the police and citizens. Several posts cannot be expected from law enforcement organisations on a daily basis; thus, there is a need for constant presence regarding special tasks related to the police like recruitment, reputation management and serving the public in a new way.

7. Validation of hypotheses

The study observed the Instagram profiles of two Hungarian law enforcement organisations, the MR and the BRFK, in the first month of their appearance on the platform, then repeated the sampling period a year later to see whether a standard and coherent organisational communication strategy could be detected. No previous study has investigated the communication strategy of Instagram profiles applied by the Hungarian law enforcement units.

After data analyses, and in response to my hypotheses, it can be stated that my first hypothesis was proved. Based on the results of the sampling periods, I found that no coherent organisational communication strategy or practice was applied by the Hungarian law enforcement units on Instagram. The content of the profiles presumably depends on the mentality and conscientiousness of the administrators, as

well as the commitment and relation of the agency's leaders to social media (Bullock 2018).

The second hypothesis was also proved by the linguistic features and vocabulary of the examined entries. The MR makes efforts and a strict commitment to engage citizens in the themes offered by Instagram. According to Bullock (2018), *'the lack of providing leadership, strategic guidance or developing an infrastructure (including a technological infrastructure) to support social media use by officers indicates that the role social media should be playing within police communications is not clear and that the degree to which utilising them is seen as an organisational priority is questionable'* (Bullock 2018, 255). It is the reason why the BRFK does not utilize the possibilities offered by the site daily, and still has room for improvement in this area, as shown by the visible contents reviewed in the sampling periods.

8. Conclusions

Regarding the interfaces of the BRFK and the MR, the study confirms that any technology can be successful and have an impact on organisational performance if it is adopted properly by organisations (Tajudeen et al. 2017). Both internationally and nationally, it can be said that to increase efficiency, this relationship must be put on a new basis, and for this purpose, the quality of external communication of the police must be improved (Peyton et al. 2019).

There were only a few posts on the profile of the BRFK; therefore, it was difficult to make a real comparison of the communication strategy of these two organisations. In the future, it is planned to extend the research to further digital platforms of these two organisations (e.g. Facebook or TikTok) or to carry out further sampling on Instagram, to study the existence of policing digilect.

The use of policing digilect as a communication practice was well received on the Instagram page of the MR and it has resulted in a positive change in the image of and opinions about the police (Uricska 2021). In the long term, it is expected that digital communication content will redefine the relationships and interrelationships among individuals, society and different organisations.

In the course of the future, citizens and organisations moving on to online platforms in the digital society and culture need to learn how to best use the strategic potential of these platforms.

References

- Babby, Earl. *A társadalomtudományi kutatás gyakorlata*. Budapest: Balassa Kiadó, 2001.
- Berelson, Bernard. *Content Analysis in Communications Research*. New York: Free Press, 1952.
- Bullock, Karen. 'The police use of social media: Transformation or normalisation?'. *Social Policy & Society* 17, no. 2 (2018): 245–58.
<https://doi.org/10.1017/S1474746417000112>

-
- Carr, Caleb T., and Rebecca A. Hayes. 'Social media: Defining, developing, and divining'. *Atlantic Journal of Communication* 23, no. 1 (2015): 46–65.
<https://doi.org/10.1080/15456870.2015.972282>
- Christián, László, Árpád Budavári, and Zsolt Lippai. 'A helyi rendészet és a településbiztonság új trendjei itthon és külföldön'. *Glossa Iuridica* 8, no. 4 (2021): 175–90.
- Fielding, Nigel G. 'Police communications and social media'. *European Journal of Criminology* (2021): 1–19.
<https://doi.org/10.1177/1477370821998969>
- Finszter, Géza. *Rendészettan*. Budapest: Dialóg Campus Kiadó, 2018.
- Hallahan, Kirk, Derina R. Holtzhausen, Betteke van Ruler, Dejan Verčič, and Krishnamurti Sriramesh. 'Defining strategic communication'. *International Journal of Strategic Communication* 1, no. 1 (2007): 3–35.
<https://doi.org/10.1080/15531180701285244>
- Heravi, Baharen Rahmanzadeh, and Natalie Harrower. 'Twitter journalism in Ireland: Sourcing and trust in the age of social media'. *Information, Communication & Society* 19, no. 9 (2016): 1194–1213.
<https://doi.org/10.1080/1369118X.2016.1187649>
- Herfort, Benjamin, Joao Porto De Albuquerque, Svend-Jonas Schelhorn, and Alexander Zipf. 'Exploring the geographic relations between social media and flood phenomena to improve situational awareness.' In *Connecting a Digital Europe through Location and Place*, edited by Joaquin Huerta, Sven Schade and Carlos Granell, 55–71. Cham: Springer, 2014.
- Highfield, Tim, and Tama Leaver. 'Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji'. *Communication Research and Practice* 2, no. 1 (2016): 1–17.
<http://doi.org/10.1080/22041451.2016.1155332>
- Istók, Béla. 'Facebook-nyelvhasználat. Normakövetés vagy funkcionális műfajmegkülönböztetés?'. In *A Selye János Egyetem X. Nemzetközi Tudományos Konferenciájának tanulmánykötete*, edited by József Bukor, Szabolcs Simon and Károly Vajda, 46–85. Komárom: Selye János Egyetem, 2018.
- Istók, Béla. 'Netnyelvészeti kaleidoszkóp'. *Eruditio – Educatio* 14, no. 1 (2019): 83–99.
- Johansen, Svein Tvedt, Bjarne Espedal, Kjell Grønhaug, and Marcus Selart. 'Envisioning trust: Trust, metaphors, and situations'. In *Proceedings of the FINT Workshop on Trust Within and Between Organizations*, 1–12. Singapore: Leibniz-Insitut für Socialwissenschaften, 2013.
<https://nbnresolving.org/urn:nbn:de:0168-ssoar-400030>
- Klaudy, Kinga. *Fordítás I: Bevezetés a fordítás elméletébe*. Budapest: Scholastica Kiadó, 1997.
- Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage, 2004.
- Kulcsár, Nárcisz. 'Motivation factors for Generation Z in higher education'. In *Developing Technology Mediation in Learning Environments*, edited by Filomena Soares, Ana Paula Lopes, Ken Brown and Anne Uukkivi, 206–20. Hershey, PA: IGI Global, 2020.
<https://doi.org/10.4018/978-1-7998-1591-4.ch012>
- Lakoff, George. 'Metaphor, Morality, and Politics, Or, Why Conservatives Have Left Liberals In the Dust.' *Social Research* 62, no. 2 (1995): 177–213.
<http://www.jstor.org/stable/40971091>

- Lee, Murray, and Alyce McGovern. *Policing and Media: Public Relations, Simulations and Communications*. London: Routledge, 2013.
- Molnár, Katalin. *Rendészeti kommunikáció – a média tükrében: Kézikönyv és szemelvénygyűjtemény a rendészeti szakújságíró szakirányú továbbképzési szak hallgatóinak*. Budapest: Dialóg Campus Kiadó, 2018.
- Peyton, Kyle, Michael Sierra-Arévalo, and David G. Rand. 'A field experiment on community policing and police legitimacy'. *Proceedings of the National Academy of Sciences* 116, no. 40 (2019): 19894–8.
<https://doi.org/10.1073/pnas.1910157116>
- Posteguillo, Santiago. 'Netlinguistics and English for internet purposes'. *Iberica* 4, no. 4 (2002): 21–38.
- Putra, Syahrul Jadmika Wignya Eka. 'Police public relations profession in the digital communication era: A phenomenological study of the adaptation of West Java police public relations personnel in the era of digital communication'. *American Journal of Humanities and Social Sciences Research* 5, no. 2 (2021): 505–10.
- Stenger, Thomas. 'Social media and online reputation management as practice: First steps towards social CRM?'. *International Journal of Technology and Human Interaction* 10, no. 4 (2014): 49–64.
<http://dx.doi.org/10.4018/ijthi.2014100104>
- Tajudeen, Farzana Parveen, Noor Ismawati Jaafar and Sulaiman Ainin. 'Understanding the impact of social media usage among organizations'. *Information & Management* 55, no. 3 (2017): 308–21.
- Uricska, Erna. 'Proper interactive communication of the police as a(n e-)trust-building strategy: Introducing the term policing digilect'. *Kosice Security Revue* 10, no. 2 (2020): 185–95.
- Uricska, Erna. 'Online reputációmenedzsment a police_hu Instagram-profilján'. *Belügyi Szemle* 69, no. 9 (2021): 1529–47.
<https://doi.org/10.38146/BSZ.2021.9.2>
- Uricska, Erna. 'The Relationship between the Framing of Speeding Messages and Changes in Attitude of Generation Z Respondents'. *Magyar Rendészet* 23, no. 1 (2023): 249–62.
<https://doi.org/10.32577/mr.2023.1.16>
- Valentine, Frances. 'Educating Gen Z in a digital world'. *Evaluation Matters – He Take Tō Te Aromatawai: Online First, New Zealand Council for Educational Research* (2019): 1–16.
- Varis, Piia. 'Digital ethnography'. In *The Routledge Handbook of Language and Digital Communication*, edited by Alexandra Georgakopoulou and Tereza Spilioti, 55–68. London: Routledge, 2016.
- Veszelszki, Ágnes. 'A digilektus hatása a dialóguslevelekre'. *Magyar Nyelv* 109, no. 4 (2013): 435–47.
- Veszelszki, Ágnes. 'Persuasion strategies on Instagram in wine communication and branding'. *Acta Universitatis Sapientiae Communicatio* 6, no. 6 (2019): 83–104.
<https://doi.org/10.2478/auscom-2019-0005>
- Veszelszki, Ágnes. *Digilect: The Impact of Infocommunication Technology on Language*. Berlin: De Gruyter Saur, 2017a.
- Veszelszki, Ágnes. *Netnyelvészet: Bevezetés az internet nyelvhasználatába*. Budapest: L'Harmattan Kiadó, 2017b.
- Winarso, Widodo. 'Organizational communication: A conceptual framework'. *Internal*

Communications & Organizational Behavior eJournal (2018).

<http://dx.doi.org/10.2139/ssrn.3171979>

Wood, Marc A., and Alyce McGovern. 'Memetic copaganda: Understanding the humorous turn in police image work'. *Crime Media Culture* 17, no. 3 (2021): 305–26.

<https://doi.org/10.1177/1741659020953452>

URL 1: roihacks.hu. 'Instagram Felhasználók száma Magyarországon (2021) – Magyar Instaszámái'. Accessed 24 February 2022.

<https://bit.ly/3xjvW7O>

URL 2: Instagram. 'brfk_budapest_police_hu'. Accessed 24 February 2021.

<https://bit.ly/3HPv8MG>

URL 3: Instagram. 'police_hu'. Accessed 24 February 2021.

<https://bit.ly/3D9oTTU>

URL 4: Instagram. 'Santa Claus on a bicycle'. Accessed 24 February 2022.

<https://bit.ly/3hAff0Q>

URL 5: Instagram. 'Kristóf Gál recites the poem "Police Work" by Belga'. Accessed 24 February 2022.

<https://bit.ly/3HItz23>

URL 6: kreativ.hu. 'Az Uniomedia lett az Év ügynöksége a Prizma 2020-on'. Accessed 3 March 2022.

<https://bit.ly/3pahAV7>