Lecture Notes in Information Systems and Organisation 68

Fred D. Davis · René Riedl ·
Jan vom Brocke · Pierre-Majorique Léger ·
Adriane B. Randolph ·
Gernot R. Müller-Putz *Editors*

Information Systems and Neuroscience

NeuroIS Retreat 2023, Vienna, Austria

Springer

Lecture Notes in Information Systems and Organisation

Volume 68

Lecture Notes in Information Systems and Organization—LNISO—is a series of scientific books that explore the current scenario of information systems, in particular IS and organization. The focus on the relationship between IT, IS and organization is the common thread of this collection, which aspires to provide scholars across the world with a point of reference and comparison in the study and research of information systems and organization. LNISO is the publication forum for the community of scholars investigating behavioral and design aspects of IS and organization. The series offers an integrated publication platform for high-quality conferences, symposia, and workshops in this field. Materials are published upon a strictly controlled double blind peer review evaluation made by selected reviewers.

LNISO is abstracted/indexed in Scopus

Fred D. Davis · René Riedl · Jan vom Brocke · Pierre-Majorique Léger · Adriane B. Randolph · Gernot R. Müller-Putz
Editors

Information Systems and Neuroscience

NeuroIS Retreat 2023, Vienna, Austria

Springer

Editors
Fred D. Davis
Information Technology Division
Texas Tech University
Lubbock, TX, USA

Jan vom Brocke
European Research Center for Information Systems (ERCIS)
University of Münster
Münster, Germany

Adriane B. Randolph
Department of Information Systems and Security
Kennesaw State University
Kennesaw, GA, USA

René Riedl
University of Applied Sciences Upper Austria
Steyr, Austria

University of Linz
Linz, Austria

Pierre-Majorique Léger
Department of Information Technologies
HEC Montréal
Montréal, QC, Canada

Gernot R. Müller-Putz
Institute of Neural Engineering
Graz University of Technology
Graz, Austria

ISSN 2195-4968 ISSN 2195-4976 (electronic)
Lecture Notes in Information Systems and Organisation
ISBN 978-3-031-58395-7 ISBN 978-3-031-58396-4 (eBook)
https://doi.org/10.1007/978-3-031-58396-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Organization

Conference Co-chairs

Fred D. Davis, Texas Tech University, Lubbock, USA
René Riedl, University of Applied Sciences Upper Austria, Steyr, Austria and Johannes Kepler University Linz, Linz, Austria

Programme Co-chairs

Jan vom Brocke, European Research Center for Information Systems (ERCIS) and University of Münster, Münster, Germany
Pierre-Majorique Léger, HEC Montréal, Montreal, Canada
Adriane B. Randolph, Kennesaw State University, Kennesaw, USA
Gernot R. Müller-Putz, Graz University of Technology, Graz, Austria

Programme Committee

Marc T. P. Adam, University of Newcastle, Callaghan, Australia
Bonnie B. Anderson, Brigham Young University, Utah, USA
Ricardo Büttner, University of Bayreuth, Bayreuth, Germany
Colin Conrad, Dalhousie University, Halifax, Canada
Alan R. Dennis, Indiana University, Indiana, USA
Thomas M. Fischer, University of Applied Sciences Upper Austria, Steyr, Austria
Rob Gleasure, Copenhagen Business School, Frederiksberg, Denmark
Jacek Gwizdka, University of Texas at Austin, Austin, Texas
Alan R. Hevner, Muma College of Business, Florida, USA
Marco Hubert, University of Aarhus, Aarhus, Denmark

Thomas Kalischko, University of Applied Sciences Upper Austria, Steyr, Austria
Peter H. Kenning, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany
C. Brock Kirwan, Brigham Young University, Utah, USA
Michael T. Knierim, Karlsruhe Institute of Technology, Karlsruhe, Germany
Jan Mendling, Humboldt-Universität zu Berlin, Berlin, Germany
Mahdi Mirhoseini, Concordia University, Montreal, Canada
Fiona F.-H. Nah, City University of Hong Kong, Kowloon Tong, Hong Kong S.A.R. (China)
Jella Pfeiffer, University of Stuttgart, Stuttgart, Germany
Sylvain Sénécal, HEC Montréal, Montréal, Canada
Fabian J. Stangl, University of Applied Sciences Upper Austria, Steyr, Austria
Stefan Tams, HEC Montréal, Montreal, Canada
Lars Taxén, Linköping University, Linköping, Sweden
Ofir Turel, California State University, California, USA
Anthony O. Vance, Virginia Tech, Virginia, USA
Eric A. Walden, Texas Tech University, Texas, USA
Peter Walla, Sigmund Freud University Vienna, Vienna, Austria
Barbara Weber, University of St. Gallen, St. Gallen, Switzerland
Selina C. Wriessnegger, Graz University of Technology, Graz, Austria

Organization Support

Fabian J. Stangl, University of Applied Sciences Upper Austria, Steyr, Austria
Thomas Kalischko, University of Applied Sciences Upper Austria, Steyr, Austria

Sponsors

We thank the sponsors of the NeuroIS Retreat 2023:

Main Sponsors

Raiffeisen Software

RAITEC IT Services

MEETING DESTINATION VIENNA

NOW ◆ TOGETHER

Technology Sponsors

BRAIN PRODUCTS
Solutions for neurophysiological research

Noldus
Information Technology

Observing the "Brain at Work" in Everyday Life Using Optical Brain Imaging: Challenges and Opportunities for NeuroIS (Keynote)

Hasan Ayaz

The understanding of the brain functioning and its utilization for real-world applications including interactions with information and communication technologies (ICTs) requires continuous, safe, accessible, mobile, and minimally intrusive monitoring. Existing studies with traditional neuroimaging approaches have accumulated overwhelming knowledge but are limited in scope, i.e. only in artificial lab settings and with simplified parametric tasks. As an interdisciplinary new field, neuroergonomics aims to fill this gap: Understanding the brain in the wild, its activity during unrestricted real-world tasks in everyday life contexts, and its relationship to action, behavior, body, and environment. The new generation of ultra-portable wearable brain imaging sensors are already positioned to go outside of the lab, and now for continuous measurements over longer periods and less constrained setups, allowing naturalistic interactions with complex ICTs. This talk will explore the latest developments in the growing research area of mobile optical brain imaging: functional near-infrared spectroscopy (fNIRS) from enabling technologies and methods to emerging field applications. As a noninvasive wearable brain-monitoring technology, fNIRS relies on optical techniques to detect changes in cortical hemodynamic responses to human perceptual, cognitive, and motor functioning. This talk will discuss emerging trends for fNIRS applications, from aerospace to medicine, with diverse populations and toward clinical solutions. Various recent synergistic fNIRS applications for human-human and human-machine interaction and interpersonal neural synchronization, highlight the potential use for NeuroIS and are ushering in the dawn of a new age in neuroscience, neuroengineering, and neuroergonomics.

Embracing Neuroadaptive Technologies: Shaping the Future of Human-Computer Interaction (Hot Topic Talk)

Thorsten O. Zander

As we progress into the future, the technologies we adopt and their applications will significantly influence the course of humanity. Effective communication between humans and technology is crucial in determining how we utilize these advancements. Brain-Computer Interfaces (BCIs) revolutionized this communication, initially enabling users to transmit direct commands without any muscular involvement. In 2011, the emergence of Passive BCIs redefined the landscape by extracting information about users' states without the need for intentional communication. A decade later, novel human-computer interaction paradigms have emerged, built upon Passive BCIs. Neuroadaptive systems, which develop an understanding of their users and autonomously adapt to their requirements, signify a convergence of human and artificial intelligence. Major corporations, militaries, governments, and startups worldwide are driving exponential growth in research and development in this domain. In this Hot Topic Talk, I will explore the potential of various neuroadaptive technologies and share examples of their early applications in human-computer interaction, artificial intelligence, and virtual reality. Predicting the legal and ethical implications of these developments on society is challenging, but I will present some considerations for discussion with the audience. Lastly, I will offer insights on how the burgeoning era of neuroadaptivity could both benefit and harm humanity's future.

Preface

The proceedings contain papers presented at the 15th annual NeuroIS Retreat held May 30–June 1, 2023. NeuroIS is a field in Information Systems (IS) that uses neuroscience and neurophysiological tools and knowledge to better understand the development, adoption, and impact of information and communication technologies (www.neurois.org).

The NeuroIS Retreat is a leading academic conference for presenting research and development projects at the nexus of IS and neurobiology. This annual conference promotes the development of the NeuroIS field with activities primarily delivered by and for academics, though works often have a professional orientation.

In 2009 the inaugural NeuroIS Retreat was held in Gmunden, Austria. Since then, the NeuroIS community has grown steadily, with subsequent annual Retreats in Gmunden from 2010 to 2017. Beginning in 2018, the conference is taking place in Vienna, Austria. Due to the Corona crisis, the organizers decided to host the NeuroIS Retreat virtually in 2020 and 2021. Starting in 2022, the NeuroIS Retreat will be held again in a physical face-to-face format in Vienna.

The NeuroIS Retreat provides a platform for scholars to discuss their studies and exchange ideas. A major goal is to provide feedback for scholars to advance their research papers toward high-quality journal publications. The organizing committee welcomes not only completed research, but also work in progress. The NeuroIS Retreat is known for its informal and constructive workshop atmosphere. Many NeuroIS presentations have evolved into publications in highly regarded academic journals.

This year is the ninth time that we publish the proceedings in the form of an edited volume. A total of 32 research papers were accepted and are published in this volume, and we observe diversity in topics, theories, methods, and tools of the contributions in this book. The 2023 keynote presentation entitled "Observing the "Brain at Work" in Everyday Life using Optical Brain Imaging: Challenges and Opportunities for NeuroIS" is given by Hasan Ayaz, associate professor at Drexel University, USA. Moreover, Thorsten O. Zander, professor for neuroadaptive human-computer interaction at the Brandenburg University of Technology in Cottbus, Germany, gives a hot topic talk entitled “Embracing Neuroadaptive Technologies: Shaping the Future of Human-Computer Interaction”.

Altogether, we are happy to see the ongoing progress in the NeuroIS field. Also, we can report that the NeuroIS Society, established in 2018 as a non-profit organization, has been developing well. We foresee a prosperous development of NeuroIS.

Lubbock, USA — Fred D. Davis
Steyr, Austria/Linz, Austria — René Riedl
Münster, Germany — Jan vom Brocke
Montréal, Canada — Pierre-Majorique Léger
Kennesaw, USA — Adriane B. Randolph
Graz, Austria — Gernot R. Müller-Putz
May 2023

Contents

Even-Related Potentials (ERPs) Reveal that Trust and Distrust Differ Between Brands and Political Institutions

Peter Walla, Stefan Kalt, and Dimitrios Külzer

Abstract Trust and distrust are important topics in the NeuroIS field. They have been described as separate constructs. This study provides evidence that brain activity differences between trust and distrust depend on stimulus category. Brand names and names of political institutions were visually presented to participants, who had to evaluate their individual trust or distrust towards those stimuli. First, it was found that stimulus category alone altered brain activities (ERPs) in frontal regions between approximately 200 and 500 ms, but also that trust and distrust differences were different between both categories. Different brain activities elicited by trusted versus distrusted political institutions occurred dominantly in the left frontal region roughly from 200 to 500 ms peaking at about 330 ms after stimulus onset. On the other hand, different brain activities elicited by trusted versus distrusted brands occurred dominantly in the right frontal region roughly from 700 to 900 ms peaking at about 780 ms after stimulus onset. These opposite hemispheric lateralizations confirm the complexity of trust and distrust. Future studies using ERPs as measures could use other stimuli (for instance related to artificial intelligence or neuroadaptive systems) relevant to the NeuroIS community and thus further our understanding of trust and distrust.

Keywords Trust · Distrust · EEG · ERPs · Brands · Political institutions

P. Walla (✉) · S. Kalt · D. Külzer
Faculty of Psychology, Freud CanBeLab, Sigmund Freud University, Freudplatz 1, 1020 Vienna, Austria
e-mail: peter.walla@sfu.ac.at

P. Walla
Faculty of Medicine, Sigmund Freud University, Freudplatz 3, 1020 Vienna, Austria

School of Psychology, Centre for Translational Neuroscience and Mental Health Research, University of Newcastle, University Drive, Callaghan, NSW, Australia

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_1

1 Introduction

Trust and distrust became very important topics on the NeuroIS research agenda [1–3]. In a review paper, three aspects of trust (genetic, hormonal and brain structures) have been summarized to present a model on the biology of trust [4, see also 5]. In the same year, another author found that trust and distrust involve different brain areas by conducting a functional Magnet Resonance Imaging (fMRI) experiment [6, see also 1 and 7]. The obvious importance of these two constructs for the NeuroIS discipline and the fact that they seem to be rather separate phenomena instead of being at opposite ends of one construct forms the basis for this study. Electroencephalography (EEG) offering an excellent temporal resolution [8, 9] has been used as well to describe trust and distrust [10], but mostly in an interpersonal context. The goal for this study was to use EEG (particularly ERPs) as objective neurophysiological measures and to focus on political institution- and brand-related trust and distrust.

For this purpose, we invited 40 participants to the Freud CanBeLab (Freud Cognitive & Affective Neuroscience and Behavior Lab) at Sigmund Freud University to record brain potential changes by using EEG, while they were presented with brand names as well as names of political institutions. Crucially, they all had to evaluate their individual trust or distrust related to all presentations and share their respective results via button presses. In order to have cleanest possible responses with respect to trust and distrust, we also offered a button to press if participants could not come to a respective conclusion and also a button if participants didn't know a particular brand or political institution.

2 Materials and Methods

We invited 40 participants, but due to heavy artifacts in 3 participants, data from only 37 were used for this study (19 males and 18 males between the ages of 18–34; mean age was 22.92 years (SD = 3.26)). All participants reported having normal or corrected-to-normal vision, being right-handed, and not having any neuropathological history. Eighty brand names (German language; e.g. Toyota, Pringles) as well as eighty names of political institutions (German language; e.g. Bauernbund, Ärztekammer) were visually presented (random order) to all participants. A 64 channel actiCHamp Plus System from Brain Products was used to record brain potential changes. PsychoPy 2021.2.3 for Windows was used to design and run the experiment. Brain potential changes were recorded with a sampling rate of 1.000 Hz (filtered: DC to 100 Hz). Offline, EEG data were down-sampled to 250 Hz and a band-pass filter from 0.1 to 30 Hz was applied. Each presentation (one trial) consisted of a 1 s long "+" symbol (fixation), a 500 ms blank screen, a stimulus (300 ms), and a 1 s blank screen. After each trial, participants were instructed to indicate via a button press, whether they felt trust or distrust towards the respective stimulus they just saw or whether they were not sure or did not even know it.

Raw EEG data were processed with EEGDISPLAY, version 6.4.9 [11]. Epochs were generated from 100 ms before stimulus onset (baseline) to 1 s after stimulus onset. Epochs with artifacts were excluded and event-related potentials (ERPs) were generated for all four conditions. All data were further down-sampled for statistical analysis (averaged across 40 ms), which was carried out by running an ANOVA (analysis of variance) followed by paired-sample t-tests. The whole study followed a $2 \times 2 \times 2 \times 4$ design. The first factor *condition* has the two levels "brand" and "political institution", the second factor *attitude* has the two levels "trust" and "distrust", the third factor *time* has the two levels "early" and "late" and the fourth factor *electrode* has the four levels "left frontal", "left parietal", "right frontal" and "right parietal".

3 Results

Visual inspection of all generated ERPs reveals a remarkably clear pattern of brain activity differences between all four conditions resulting in two main findings. First, both stimulus categories elicited remarkably different brain activities in frontal brain areas roughly from 200 to 500 ms after stimulus onset (Fig. 1). Tables 2 and 3 present t-test results that reveal this difference to be significant for a representative time point at 330 ms (Table 2), but not for the representative later time point at 780 ms (Table 3). Second, brain activity differences between trust and distrust varied depending on stimulus category, i.e. brands or political institutions. Analysis of variance (ANOVA) reveals a significant 4-way interaction between all factors ($F(1,746) = 4{,}175$, $p = 0.024$, $\eta 2 = 0.104$) (Table 1) proving this finding. T-test results presented in Table 3 show that brain activities elicited by trusted political institutions differ from brain activities elicited by distrusted political institutions at the early time point (330 ms after stimulus onset) at electrode position F7. T-test results also support that brain activities elicited by trusted brands differ significantly from brain activities elicited by distrusted brands at the later time point (780 ms after stimulus onset) at electrode position F8 (Table 2). Figure 2 shows topographical maps underlying all those findings.

4 Discussion

As has been well summarized and mentioned in two NeuroIS papers, one published in 2015 [9] and the other in 2020 [3], EEG and also calculated ERPs from raw EEG data represent a valuable tool in NeuroIS research. In the earlier paper [9], it is also highlighted and explained that ERPs are inevitably associated with very distinct stimuli that must at least have a clear onset in terms of exposing them to one or more sensory systems of a study participant. Since our interest was in trust versus distrust (two highly relevant topics in NeuroIS research; [1, 4, 6, 7]) in brand names

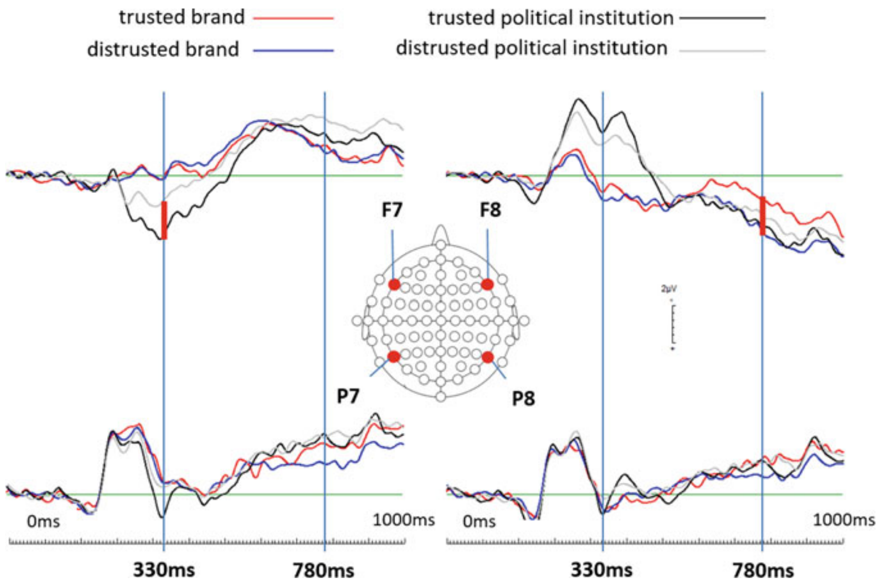


Fig. 1 ERPs of all four conditions. At F7, ERPs from trusted and distrusted political institutions differ from each other (highest statistical significance at 330 ms after stimulus onset). ERPs from trusted and distrusted brands differ from each other at F8 (most significantly at 780 ms after stimulus onset). At both posterior electrode locations P7 and P8 no significant differences occurred

Table 1 Analysis of variance (ANOVA) results

Attitude main effect and interactions	df	F	p-value	η2
Attitude	1000	0.013	0.911	0.000
condition * attitude	1000	3.538	0.068	0.089
time * attitude	1000	2.657	0.112	0.069
condition * time * attitude	1000	0.230	0.635	0.006
electrode * attitude	2350	5.605	0.003	0.135
condition * electrode * attitude	1642	0.613	0.514	0.017
time * electrode * attitude	1975	1.416	0.249	0.038
condition * time * electrode * attitude	**1746**	**4.175**	**0.024**	**0.104**

and names of political institutions (both distinct stimuli) and in recording respective brain activities with high temporal resolution, it was clear that ERPs are the best approach to this.

Our first finding that brand names and names of political institutions elicit significantly different ERPs might seem rather trivial, but the fact that ERPs from individual trust and distrust evaluations related to these two word categories differ in category-specific ways is interesting and potentially provides further help to understand the constructs of trust and distrust.

Table 2 T-test results for comparisons between brain activities elicited by trusted brands and distrusted brands as well as between trusted and distrusted political institutions (p. I.) at 330 ms after stimulus onset at left frontal electrode location F7

T-test comparisons (F7; 330 ms)	SD	T	df	p
trusted brand – distrusted brand	2389	−0.191	36	0.849
trusted brand – trusted p. I	3388	−5.662	36	0.000
trusted brand – distrusted p. I	2801	−2.969	36	0.005
distrusted brand – trusted p. I	2682	−6.981	36	0.000
distrusted brand – distrusted p. I	3011	−2.610	36	0.013
trusted p. I. – distrusted p. I	**3591**	**3.026**	**36**	**0.005**

Table 3 T-test results for comparisons between brain activities elicited by trusted brands and distrusted brands as well as between trusted and distrusted political institutions (p. I.) at 780 ms after stimulus onset at right frontal electrode location F8

T-test comparison (F8; 780 ms)	SD	T	df	p-value
trusted brand – distrusted brand	**4633**	**−2200**	**36**	**034**
trusted brand – trusted p. I	4805	−1699	36	098
trusted brand – distrusted p. I	3957	−1282	36	208
distrusted brand – trusted p. I	2892	701	36	488
distrusted brand – distrusted p. I	2752	1859	36	071
trusted p. I. – distrusted p. I	2552	1211	36	234

Brand-related trust and distrust elicited differing ERPs in right frontal cortical regions in a time-window roughly spanning from 700 to 900 ms after stimulus onset. On the other hand, political institution-related trust and distrust elicited differing ERPs in mainly left frontal cortical regions in a time window roughly spanning from 200 to 500 ms after stimulus onset. Given the fact that hemispheric functional differences exist, this finding might lead to the notion that trust and distrust evaluations (or in other words attitudes towards both word categories) are related to aspects that differ between brands versus political institutions. The right hemisphere is synthetic, creative and emotional [12, 13], while the left hemisphere is analytic, logic and language-based [13, 14]. Furthermore, lateralization of trust and distrust in general supports fMRI findings and transactional, theoretical designs [15].

In light of the present findings, future research using the ERP approach could include other stimuli such as for instance artificial intelligence-related content like descriptions of speech assistants or text-based chatbots like ChatGPT. This seems particulary interesting to the NeuroIS community since the effect of personality traits on trust in artificial intelligence has already been summarized [16].

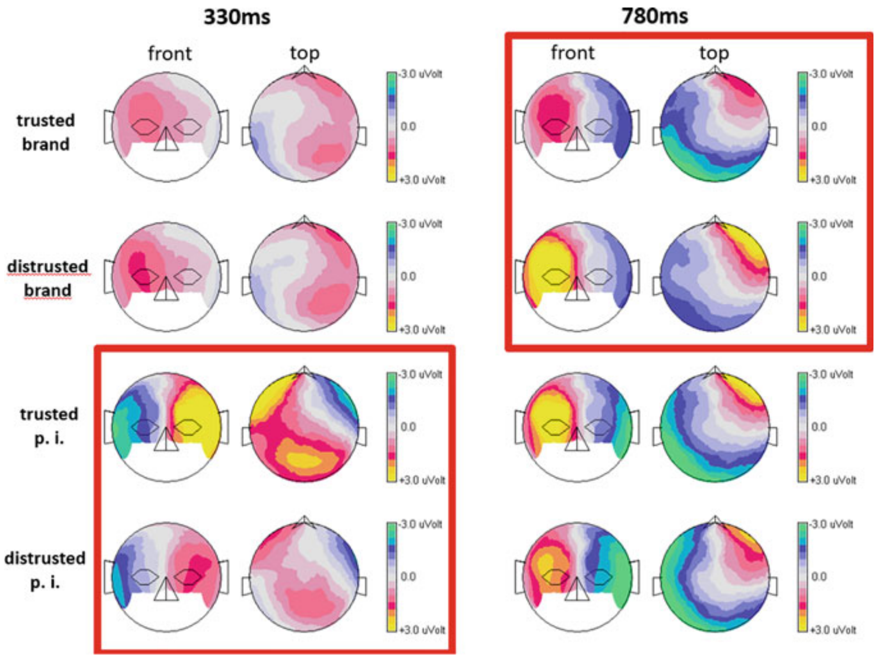


Fig. 2 Topographical maps: Brain activities differ between trusted and distrusted political institutions (p.i.) at 330 ms and between trusted and distrusted brands maijnly at 780 ms after stimulus onset

References

1. Riedl, R., Hubert, M., & Kenning, P. (2010). Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *MIS Quarterly, 34*(2), 397–428.
2. vom Brocke, J., Hevner, A., Majorique Léger, P., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*(1), 9–24.
3. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020, August). A decade of NeuroIS research: Progress, challenges, and future directions. *SIGMIS Database, 51*(3), 13–54.
4. Riedl, R., & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 63–91.
5. Krueger, F., & Meyer-Lindenberg, A. (2019). Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends in Neurosciences, 42*(2), 92–101.
6. Dimoka, A. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly, 34*(2), 373–396.
7. Riedl, R., Mohr, P., Kenning, P., Davis, F., & Heekeren, H. (2014). Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems, 30*(4), 83–114.
8. Walla, P., Georg, N., & Herbert, C. (2021). The human self has two serial aspects and is dynamic: A concept based on neurophysiological evidence supporting a multiple aspects self theory (MAST). *Life, 11*(7), 611.

9. Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems, 37*.
10. Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience, 4*(1), 23–34.
11. Fulham, W. R. (2015). *EEG display (Version 6.4.9) [Computer software].* Functional Neuroimaging Laboratory, University of Newcastle.
12. Lindell, A. K. (2013). Continuities in emotion lateralization in human and nonhuman primates. *Frontiers in Human Neuroscience, 7*, 464.
13. Corballis, M. C. (2014). Left brain, right brain: Facts and fantasies. *PLOS Biology, 12*(1), 1–6.
14. Walla, P., Hufnagl, B., Lindinger, G., Imhof, H., Deecke, L., & Lang, W. (2001). Left temporal and temporoparietal brain activity depends on depth of word encoding: A magnetoencephalographic study in healthy young subjects. *NeuroImage, 13*(3), 402–409.
15. Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*(3), 277–283.
16. Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electron Markets, 32*, 2021–2051.

RACE: A Real-Time Architecture for Cognitive State Estimation, Development Overview and Study in Progress

Noémie Beauchemin, Alexander John Karran, Jared Boasen, Bella Tadson, Patrick Charland, François Courtemanche, Sylvain Sénécal, and Pierre-Majorique Léger

Abstract Cognitive load management is important in successful learning, referring to working memory and other factors related to accomplishing instructional tasks. Cognitive overload and underload are induced when challenges provided to the student exceed or underutilize working memory capacity, leading to suboptimal learning. The link between cognitive load and successful learning is well established. However, current educational technologies fail to utilize cognitive load effectively to personalize learning and fail to adapt to the student's learning pace. Neuroadaptive interfaces, specifically Brain-Computer Interfaces, are slowly transforming the traditional educational landscape offering promising possibilities to enhance and improve learning experiences by enabling direct communication between the brain and a computer to adapt instructional content in real-time based on the assessment of cognitive load brain states. This research-in-progress paper discusses the development, following a design science research methodology, of *RACE*: a novel artefact consisting of a Closed-Loop Brain-Computer Interface that measures cognitive load in real-time applied to a memorization-based learning task to adapt the learning Interactive User Interface in real-time based on assessed and classified levels of cognitive load. Specifically, this artefact adapts the speed of information provision and response time to the learner's pace to make learning more personalized and effective.

N. Beauchemin (✉) · A. J. Karran · J. Boasen · B. Tadson · F. Courtemanche · S. Sénécal · P.-M. Léger (✉)
Tech3Lab, HEC Montréal, Montréal, Canada
e-mail: noemie.beauchemin@hec.ca

P.-M. Léger
e-mail: pml@hec.ca

J. Boasen
Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

P. Charland
Département de Didactique, Université du Québec à Montréal, Montréal, Canada

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_2

Keywords Neuroadaptive interface · Brain-computer interface · Biocybernetic loop · Learning · Cognitive load theory · Design science

1 Introduction

The development and integration of technologies into teaching practices have begun a trend toward transitioning from the more traditional classroom pedagogical models to online models [1, 2]. Research has shown that the use of technological tools in learning helps promote engagement and motivation as predictors of success [1, 3]. While technological tools in education have been designed with user Cognitive Load (CL) as a design consideration [4, 5] very few of these technologies utilize direct real-time measurements of CL to adapt in real-time, potentially making learning less personalized [6].

Neuroadaptive technologies, specifically Brain-Computer Interfaces (BCI), are tools to overcome physical impairments and augment specific cognitive capacities [7]. Rapid improvements in sensor technologies such as electroencephalogram (EEG) and methods of classifying brain activity into specific states have shown BCI to be a useful assistive and interfacing technology [38] for human–machine systems [8]. BCI technology has been defined as "*a device that reads voluntary changes in brain activity, then translates these signals into a message or command in real-time*" [9]. BCIs are a core component of systems that utilize the user's neurophysiological data as input to a computer system, which then performs actions to adapt, assist or provide feedback to the operator. A common application of BCI technology is to measure and classify CL under various conditions. Studies have found correlations between CL and variance in brainwaves expressed as increases or decreases in α (8–12 Hz) and θ (4–8 Hz) in pre-frontal brain regions [10].

In this work-in-progress manuscript, we answer a call for research to investigate neuroadaptive technology using NeuroIS methods [11, 12] and discuss the integration of a design science approach to developing a research BCI artefact that monitors and classifies CL in real-time to drive interface adaptions to improve learning outcomes in an education context. We provide an overview of the requirements analysis, design choices and overall architecture of the BCI artefact and provide a study methodology that utilizes the BCI artefact to adapt an interface in two ways: speed of information presentation and response time, to investigate if these adaptations improve learning outcomes.

As the means from which to derive requirements for the BCI artefact that meet the needs of the study, we posit the following research question, "*To what extent does utilizing a real-time BCI that adapts the speed of information provision and response times based on cognitive load improve learning outcomes in a task involving memorization of astronomical constellations?*".

2 Background

Cognitive Load and Learning

Many factors influence learning; however, CL remains a central concept for understanding and improving the learning process [5, 13, 14]. Cognitive Load Theory (CLT), proposed by Sweller [15], posits a cognitive architecture to investigate how information is processed and retained and centers around the interactions between Working Memory (WM) and long-term memory [6, 14–16]. It defines CL as the management of the WM's limited capacity, i.e. the amount of mental effort an individual allocates to a task [6].

Cognitive overload or underload during the completion of online or computer-based learning tasks may occur when WM's capacity is exceeded or underutilized, potentially leading to slow learning progress or poor performance [17]. Current educational technologies consider CL as only one of many factors influencing learning outcomes and do not emphasize its centrality to the learning process or how modulating CL may lead to improved learning outcomes [4, 5].

Previous methods of quantifying CL in both research and developing educational technologies consisted of batteries of subjective measures administered through questionnaires [14, 18]. However, while these measures provide the learner's perspective on their experience, they cannot quantify the amount of mental effort invested throughout the entire learning process [14]. One solution to this problem is to measure CL directly and in real-time through the brain's electrical activity using BCI.

Brain-Computer Interfaces

As discussed previously, BCIs are systems that allow the human brain to communicate directly with a computer [19]. BCIs transform brain activity into control signal data for computer interaction [20, 21]. BCI research has gained in popularity in the last decade due to its potential clinical application [20]. These systems allow bypassing the peripheral nervous system for neurorehabilitation in cases of brain injury, motor disabilities and other medical purposes [19, 22, 23]. BCI technology has also been used in studies investigating video games [24–29], marketing and advertisement [30, 31], neuroergonomics and smart environments [32–37], and work monitoring and safety [38–43]. There are currently three categories of BCI: *Active*, where users voluntarily and consciously control their brain activity to directly control an application [8, 25]; *Reactive*, a hybrid of *Active* and *Passive* paradigms, where users indirectly modulate their brain activity in response to external stimuli, using Event-Related Potentials (ERPs) derived from brain activity, to control an application [8, 25]; and *Passive*, wherein spontaneous brain activity is automatically monitored to differentiate or quantify mental states, where the user provides no active control and where feedback is provided as a response from the system [8, 25, 44].

Interest in neurotechnology and more specifically passive BCIs has grown rapidly in the last decade [45]. In a passive BCI, brain activity is classified, then these classifications are sent to a computer system, which then adapts content or provides visual feedback, which in turn encourages changes in brain activity as part of a biocybernetic loop [46]. There are several examples of passive BCIs in the literature which have been used to support learning tasks [47], increase engagement [48], and increase performance [49] of learners.

While interest in BCIs has grown substantially, few research papers exist regarding BCI technologies focused on learning and measuring learner's CL in real-time. Furthermore, while the theoretical relationship between learning and cognitive load is strong, and several research studies [40–42, 47, 48] have been conducted to develop BCIs to detect levels of CL, none specifically focuses on utilizing CLT and BCI technology to monitor CL and adapt learning content to the user in real-time.

Speed of Stimulus Presentation in Learning

Learning pace, modulated by the speed of stimulus presentation, has been extensively studied for decades [50]. The need to adapt, personalize and present content to the learner's pace to increase information retention and improve learning has been noted many times [51–53]. In this context, a BCI could be utilized to monitor CL in real-time and trigger an interface to adapt and personalize the pace of learning. Most previous research using BCIs in an educational context applied the technology to assess mental state concerning interface complexity and CL while using a new interface and not directly adapting learning content [6, 14, 44, 50]. To our knowledge, the research presented here is the first of its kind proof of principle as it integrates BCI technology, real-time measurement of CL and speed of stimulus presentation to create a neuro-adaptive learning interface. It is, therefore, imperative to follow a rigorous *Design Science Research Methodology* to develop a complete and valid solution.

3 Objectives and Methodology

We created our neuro-adaptive artefact in accordance with Brocke et al. [54] and following Peffers et al.'s [55] Design Science Research Methodology (DSRM). The DSRM provides a valuable framework for our research use case, given its wide adoption [56] and iterative nature [55]. First, we formulated a problem statement: "*design an artefact that can regulate the level of cognitive load of users while performing a learning task*". Second, we performed a series of iterative development activities (*Activity 1–6*) to develop a valid artefact.

We began our methodological process with *Activity 1*, which consisted of an in-depth analysis of the current literature concerning our research problem: the absence of a reliable and valid system in the field of education to regulate the cognitive load of learners to improve their learning. This analysis was necessary to fully understand all aspects of the problem and to create a relevant and useful solution. Theoretical foundations were drawn from previous research on CL and BCIs (see other sections on CL and BCI), and were applied to our design.

In *Activity 2*, we explored the state of existing and potential solutions and formulated objectives (see next section) that could help solve the identified problem. To build our objectives, we examined the rigor of the different methodologies used in the previous research, thereby following a rigor and relevance process [55, 57]. Since there are very few studies about BCIs and learning, objectives were aligned with a *Type I* use case, which centers the BCI as a tool for research purposes [58].

Subsequently, we proceeded with the Design Cycle throughout *Activity 3*. We developed the solution following an iterative process through several research activities and design-related decisions until the solution fulfilled its objectives extended over an 8-month period. Specifically, we have conducted 12 main research activities related to the IUI and the neuroadaptive system through just over 50 pre-tests, resulting in approximately 45 design-related decisions and iterations.

We then continued with *Activity 4*, which allowed us to demonstrate with a small sample of participants that the artifact does indeed adapt in real-time according to a classification of CL, therefore confirming its feasibility and practical potential. We were able to test the solution on 10 pre-test participants.

Afterwards, we assessed the quality and validity of the artifact through simulations to demonstrate that (1) the adaptations occurred as expected and (2) that it met all the initial design requirements as part of *Activity 5*. In future steps, we plan to test the artifact in larger-scale controlled experiments to assess its performance and effect on cognitive workload in a learning context. We also plan to communicate our DR and results to the scientific community through publication as part of *Activity 6*.

To achieve our goal and cover the broadest range of features required to fulfill a functional BCI artifact, we derived a series of four design objectives (DO).

DO1: *The interactive user interface (IUI) should support a learning task which displays an image of a star constellation with associated multiple-choice answers and capture feedback (as right or wrong answers) for a predetermined amount of time adapting to a user's level of cognitive load.* To create the learning task, we adapted Riopel et al.'s [59] constellation memorization study to create a valid task capable of inducing CL fluctuations. However, for this study, we selected 32 constellations based on unfamiliar names or confusingly similar visual forms (see Fig. 1). In our study, the adaptive parameter influenced by the user's CL is the speed of information provision, more precisely (1) the amount of time given to answer and (2) the amount of time for the answer feedback. Both should have the same duration and change on the IUI according to the level of the CL classifier. Right or wrong answers should not affect the speed of information provision. Thereby, the IUI should permit isolation of the effect of the speed of information provision to adequately measure the CL. According to the current literature on CL, the IUI should be as clear as possible by avoiding too many different elements (figures, colors, etc.) and redundant text to minimize extraneous processing and by avoiding complex sentences to minimize

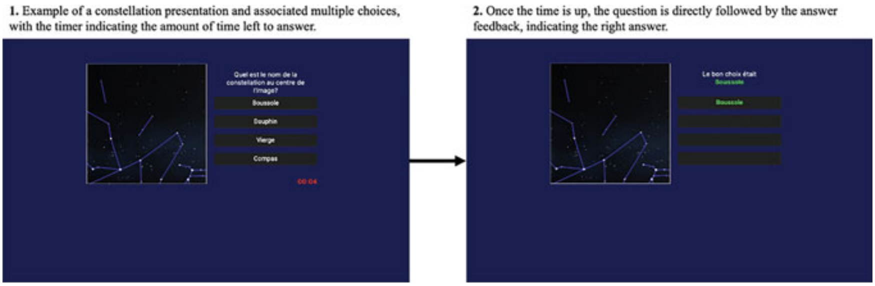


Fig. 1 Design of the Interactive User Interface (IUI) following the Design Objective 1 (DO1) and 2 (DO2)

intrinsic processing and maximize germane processing [60]. The IUI should also show the right answer even when the user answers correctly. Previous research shows that even if a right answer is obtained, following feedback is important for better information retention and to avoid making future mistakes [61, 62]. Finally, the task duration should be long enough to ensure CL fluctuations over time.

DO2: *The system should regulate cognitive load levels through neurofeedback by adapting the information presentation speed of an interface (i.e., stimulus speed of presentation) to improve users learning and enhance their performance.* The adaptation should not obstruct the learning task itself. Therefore, the IUI informs the user of how many seconds are left to answer the question with a countdown timer right underneath the multiple-choice answers (see Fig. 1). The countdown timer should be displayed in a way that is easily perceived by the user without creating anxiety or stress and without affecting recall performances [63]. Changes in the speed of information provision have to be relatively subtle to not interfere with the task and performances, but relevant enough to create a brain state change in the user. Thus, the amount of time given to answer the question and the amount of time for the answer feedback both increase or decrease with 1 s jumps at a time, going as high as 8 s and as low as 3 s each. The minimum was set at 3 s to avoid to avoid transient brain responses to novel information being confounded with CL classification. The maximum was established based on pretests and observations of time limits where participants begin to disengage with the task.

DO3: *The system should classify the level of cognitive load continuously and in real-time and communicate the level of cognitive load to the IUI.* To fulfill this requirement, we used a Lab Streaming Layer (LSL) to communicate CL classifiers to a Python script that sends the classifiers to the IUI through a Web Socket client. Classifiers were transmitted from the start to the end of the experiment every six seconds.

DO4: *The system should record and store raw neurophysiological data during use for post-hoc analysis.*

4 Design and Development

Interactive User Interface and Adaptation Logic

Figure 2 illustrates the proposed artefact's process flow, which follows the four design objectives and iterative design activities. The artefact was developed in Simulink MATLAB (version R2021b, Mathworks MA) and uses a wireless 32-channel active electrode EEG from G.Tec (g.Nautilus, Austria) to continuously measure brain activity. To act as a baseline for post-hoc analysis, a small and static black square in the middle of a gray screen for 1 min and 30 s was displayed before the experiment began. To train the artefact and set threshold values for a high and low workload classification, we developed an *n*-back task where $n = 0$ and $n = 2$. Used in many studies to induce high (2) and low (0) CL through the manipulation of WM [10, 64–67], the *n*-back task was deemed to be the most appropriate calibration task for CL classification because it requires the memorization and recall of presented visual stimuli, similar to the constellation learning task.

To support the instantiation of **DO3**, the artefact processes end-to-end the acquired brain signals and classifies CL as low (0), medium (1), and high (2) through a novel index calculation based on mean alpha band power in the parietal cortex over a 6 s sliding window, stabilized by comparing average CL calculated using a sliding window of 60 s. Classifications are sent via Lab Streaming Layer (LSL) to a Python script which then pushes the level of CL to the Interactive User Interface (IUI) every 6 s through a WebSocket client integrated into a dynamic Web app built with AngularJS. We implemented a rule "engine" to allow the web app to switch from active (experimental) to passive (control) conditions, whereby the neuro-adaptivity rules are provided through a JSON file on selecting "active". When either option is selected a personalized link is generated leading to the correct IUI for each participant, further generating placeholder database entries to store the behavioral and qualitative

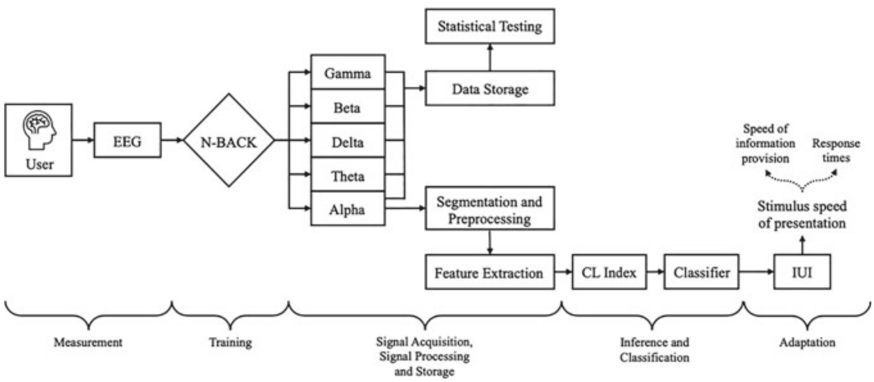


Fig. 2 Real-time Architecture for Cognitive State Estimation (RACE), Block process diagram of the BCI system, moving from User to the IUI

data for later analyses. The IUI was presented to the participants via Google Chrome. The IUI displays a constellation image and four multiple-choice questions with a timer indicating the remaining response time, followed by the correct answer (see Fig. 1).

The neuro-adaptivity model is integrated into the IUI; when CL is high (2), the artefact decreases the speed of information provision and increases response time each by one second (max 8 s). When CL is low (0), the artefact increases the speed of information provision and decreases response time by one second each (min 3 s). Starting time is set at 5 s, and no adaptation occurs when the IUI receives a “1”.

5 Next Steps: Artifact Evaluation and Experimental Study

We have evaluated the artefact through pre-tests and confirmed that its development meets all the initial design objectives, demonstrates a high level of utility in learning, and has the potential to go beyond the boundaries of research and laboratory application [55, 56, 68]. Our next step is to evaluate the artefact in a controlled laboratory study with a larger pool of participants. To this end, we developed a between-subjects study design to isolate the effect of neuro-adaptivity. In group one (control), the speed of information provision is the same throughout each trial block (without neuro-adaptivity); in group two (experimental), the speed of information provision varies according to the participant’s cognitive load level (neuro-adaptivity). The task involves learning and memorizing as many constellations as possible from a total of 32 constellations. The task consists of four trial blocks, separated by a 30 s break, where each constellation is presented two times per trial block. As per design specification, multiple-choice answers are randomly presented, and the correct answer’s position between all four possible answers is also randomized. The presentation order of the constellations in each trial block has been pre-randomized and is identical for all participants. We evaluate participant performance throughout the experiment. Before the experiment begins, participants are asked to complete a short questionnaire including the 10-item *Edinburgh Handedness Inventory* to assess handedness [69], demographic questions and questions about prior level of interest and knowledge of constellations. A second short questionnaire is presented to the participants immediately after the experiment to gather self-repoted data on their experience, including the *NASA-TLX* to estimate perceived workload [70], the *System Usability Scale* (SUS) to measure the perceived usability of the system [71] and the 5 dimensions of Cognitive Absorption (Temporal Dissociation, Focused Immersion, Heightened Enjoyment, Curiosity and Control) of the *Psychologigal Ownership of IT* (POIT) [72]. The study is currently in progress, we have gathered data for n = 45 participants for evaluation and statistical testing and we look forward to sharing our preliminary results.

References

1. Bergdahl, N., Nouri, J., & Fors, U. (2020). Disengagement, engagement and digital skills in technology-enhanced learning. *Education and Information Technologies, 25*(2), 957–983.
2. Alharthi, M. (2020). Students' attitudes toward the use of technology in online courses. *International Journal of Technology in Education, 3*(1), 14–23.
3. Fırat, M., Kılınç, H., & Yüzer, T. V. (2018). Level of intrinsic motivation of distance education students in e-learning environments. *Journal of Computer Assisted Learning, 34*(1), 63–70.
4. Gerjets, P., et al. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience*, 8.
5. Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1–16.
6. Kalyuga, S., & Liu, T.-C. (2015). Guest editorial: Managing cognitive load in technology-based learning environments. *Journal of Educational Technology & Society, 18*(4), 1–8.
7. Aggarwal, S., & Chugh, N. (2022). Review of machine learning techniques for EEG based brain computer interface. *Archives of Computational Methods in Engineering*, 1–20.
8. Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering, 8*(2), 025005.
9. Guger, C., Allison, B. Z., & Gunduz, A. (2021). Brain-computer interface research: A state-of-the-art summary 10. *Brain-computer interface research* (pp. 1–11). Springer.
10. Grimes, D., et al. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*
11. Riedl, R., & Léger, P. (2016). Fundamentals of NeuroIS: Studies in neuroscience. In *Psychology and behavioral economics.* Springer.
12. vom Brocke, J., et al. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*(1), 9–24.
13. van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*(3), 5–13.
14. Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education, 141*, 103618.
15. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.
16. Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.
17. Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.
18. Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53–61.
19. Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-computer interfaces in medicine. *Mayo Clinic Proceedings, 87*(3), 268–279.
20. Lotte, F., et al. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update. *Journal of Neural Engineering, 15*(3), 031005.
21. Dimoka, A., et al. (2012). On the use of neurophysiological tools in is research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702.
22. Abiri, R., et al. (2019). A comprehensive review of EEG-based brain–computer interface paradigms. *Journal of Neural Engineering, 16*(1), 011001.
23. Chaudhary, U., Birbaumer, N., & Ramos-Murguialday, A. (2016). Brain–computer interfaces for communication and rehabilitation. *Nature Reviews Neurology, 12*(9), 513–525.

24. Lalor, E. C., et al. (2005). Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment. *EURASIP Journal on Advances in Signal Processing, 2005*(19), 706906.
25. Ahn, M., et al. (2014). A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors, 14*(8), 14601–14633.
26. van de Laar, B., et al. (2013) Experiencing BCI control in a popular computer game. *IEEE Transactions on Computational Intelligence and AI in Games, 5*(2), 176–184.
27. Kerous, B., Skola, F., & Liarokapis, F. (2018). EEG-based BCI and video games: A progress report. *Virtual Reality, 22*(2), 119–135.
28. Lécuyer, A., et al. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer, 41*(10), 66–72.
29. Labonte-Lemoyne, E., et al. (2018). Dynamic threshold selection for a biocybernetic loop in an adaptive video game context. *Frontiers in Human Neuroscience*, 12.
30. Mashrur, F. R., et al. (2022). BCI-Based Consumers' choice prediction from EEG signals: An intelligent neuromarketing framework. *Frontiers in Human Neuroscience, 16*, 861270.
31. Bonaci, T., Calo, R., & Chizeck, H. J. (2015). App stores for the brain: privacy and security in brain-computer interfaces. *IEEE Technology and Society Magazine, 34*(2), 32–39.
32. Abdulkader, S. N., Atia, A., & Mostafa, M.-S.M. (2015). Brain computer interfacing: Applications and challenges. *Egyptian Informatics Journal, 16*(2), 213–230.
33. Lin, C. T., et al. (2014). Brain computer interface-based smart living environmental auto-adjustment control system in UPnP home networking. *IEEE Systems Journal, 8*(2), 363–370.
34. Kosmyna, N., et al. (2016). Feasibility of BCI control in a realistic smart home environment. Frontiers *in Human Neuroscience*, 10.
35. Tang, J., et al. (2018). Towards BCI-actuated smart wheelchair system. *BioMedical Engineering OnLine, 17*(1), 111.
36. Carabalona, R., et al. (2012). Light on! Real world evaluation of a P300-based brain–computer interface (BCI) for environment control in a smart home. *Ergonomics, 55*(5), 552–563.
37. Wascher, E., et al. (2023). Neuroergonomics on the go: An evaluation of the potential of mobile EEG for workplace assessment and design. *Human Factors, 65*(1), 86–106.
38. Venthur, B., et al. (2010). Novel applications of BCI technology: Psychophysiological optimization of working conditions in industry. In *2010 IEEE International Conference on Systems, Man and Cybernetics.*
39. Roy, R. N., et al. (2013). Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).*
40. Karran, A. J., et al. (2019). Toward a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. *Frontiers in Human Neuroscience,* 13.
41. Demazure, T., et al. (2021). Enhancing sustained attention. *Business & Information Systems Engineering, 63*(6), 653–668.
42. Demazure, T., et al. (2019). Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface. In *Information systems and neuroscience*. Springer International Publishing.
43. Knierim, M. T., Zimny, C. & Seitz, J. (2022). Monitoring cognitive load with 3D-printed EEG headphones.
44. Krol, L. R., & Zander, T. O. (2017). Passive BCI-based neuroadaptive systems. In *GBCIC*.
45. Aricò, P., et al. (2018). Passive BCI beyond the lab: Current trends and future directions. *Physiological Measurement, 39*(8), 08TR02.
46. Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology, 40*(1–2), 187–195.
47. Andreessen, L. M., et al. (2021). Toward neuroadaptive support technologies for improving digital reading: A passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. *User Modeling and User-Adapted Interaction, 31*(1), 75–104.

48. Eldenfria, A., & Al-Samarraie, H. (2019). Towards an online continuous adaptation mechanism (OCAM) for enhanced engagement: An EEG study. *International Journal of Human-Computer Interaction, 35*(20), 1960–1974.
49. Yuksel, B. F., et al. (2016). Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.*
50. Petko, D., Schmid, R., & Cantieni, A. (2020). Pacing in serious games: Exploring the effects of presentation speed on cognitive load, engagement and learning gains. *Simulation & Gaming, 51*(2), 258–279.
51. Najjar, L. J. (1996). Multimedia information and learning. *Journal of Educational Multimedia and Hypermedia, 5*(2), 129–150.
52. O'Byrne, W. I., & Pytash, K. E. (2015). Hybrid and blended learning: Modifying pedagogy across path, pace, time, and place. *Journal of Adolescent & Adult Literacy, 59*(2), 137–140.
53. Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments, 7*(1), 33.
54. Brocke, J. V., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems, 53*(3), 1–13.
55. Peffers, K., et al. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*, 45–77.
56. vom Brocke, J., Hevner, A., & Maedche, A. (2020). *Introduction to design science research* (pp. 1–13).
57. Hevner, A. R., et al. (2004). Design science in information systems research. *MIS Quarterly,* 75–105.
58. Blankertz, B., et al. (2016). The berlin brain-computer interface: Progress beyond communication and Control. *Frontiers in Neuroscience*, 10.
59. Riopel, M., et al. (2017). Using invariance to model practice, forgetting, and spacing effects. In *EDULEARN17 Proceedings.* IATED.
60. DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*(1), 223.
61. Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), 211–232.
62. Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 918–928.
63. Ghafurian, M., Reitter, D., & Ritter, F. E. (2020). Countdown timer speed: A trade-off between delay duration perception and recall. *ACM Transactions on Computer-Human Interaction (TOCHI), 27*(2), 1–25.
64. Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience, 8*, 322.
65. Brouwer, A.-M., et al. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering, 9*(4), 045008.
66. Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using wireless EEG signals to assess memory workload in the n-back task. *IEEE Transactions on Human-Machine Systems, 46*(3), 424–435.
67. Jansma, J. M., et al. (2000). Specific versus nonspecific brain activity in a parametric N-back task. *NeuroImage, 12*(6), 688–697.
68. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–355.
69. Caplan, B., & Mendoza, J. E. (2011). Edinburgh handedness inventory. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 928–928). Springer.

70. Hart, S. G., & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139–183). Elsevier.
71. Brooke, J. (1996). SUS—A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4–7.
72. Barki, H., Pare, G., & Sicotte, C. (2008). Linking IT implementation and acceptance via the construct of psychological ownership of information technology. *Journal of Information Technology, 23*(4), 269–280.

Exploring the Role of Post-hoc Explanations in Mitigating Algorithm Aversion in Identity-Based Consumption: An Eye-Tracking Study

Yannik Schlepper, Bernhard Lutz, Jörg Lindenmeier, and Dirk Neumann

Abstract Customers have a general tendency to discount algorithmic over human recommendations, a phenomenon commonly known as "algorithm aversion." Within areas driven by identity-based consumption such as fashion, designing efficient recommender systems is particularly challenging due to highly individualistic preferences and tastes. In this study, we analyze algorithm aversion towards fashion recommender systems with regards to social and personal identity and post-hoc explanations of algorithmic recommendations. In line with self-categorization theory and theory of planned behavior, we hypothesize that, to minimize algorithm aversion, the post-hoc explanations of algorithmic recommendations need to target customers' salient identity. Accordingly, we propose a 3 × 3 between-subject experiment with eye-tracking, where participants are shown several pairs of algorithm- or human-based fashion recommendations. In the treatment groups, we either activate customers' social or personal identity, while the explanations of algorithmic recommendations emphasize the customers' mainstream or unique taste. Furthermore, we expect that consumers with activated social or personal identity are more likely to report a different preference than their preference measured by the first and total number of eye fixations. Thereby, we expect to extend IS research on algorithm aversion and post-hoc explanations of algorithms towards identity-based consumption. In addition, our findings have practical implications for online retailers.

Keywords Algorithm aversion · Identity-based consumption · Post-hoc explanations · Eye-tracking · Recommender systems

Y. Schlepper (✉) · B. Lutz · J. Lindenmeier · D. Neumann
University of Freiburg, Freiburg, Germany
e-mail: yannik.schlepper@wvf.uni-freiburg.de; yannik.schlepper@gmail.com

B. Lutz
e-mail: bernhard.lutz@is.uni-freiburg.de

J. Lindenmeier
e-mail: joerg.lindenmeier@vwl.uni-freiburg.de

D. Neumann
e-mail: dirk.neumann@is.uni-freiburg.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_3

1 Introduction

Humans tend to discount algorithmic over human recommendations—even when the algorithm is superior, a phenomenon called "algorithm aversion" [1, 2]. Past research has explored various drivers of algorithm aversion in the context of objective high-stakes decisions like financial investments [3] or medical diagnosis [4]. Possible reasons of algorithm aversion are a general lack of trust in technology, a desire to maintain control, or a missing understanding of how the algorithm works [5–7]. However, algorithm aversion also occurs in the context of subjective tasks as humans tend to believe that algorithms lack the human capabilities of being creative, or account for unique situations and preferences [7–10].

The lack of human capabilities provides additional challenges for IS research and practice in mitigating algorithm aversion towards recommender systems in identity-based consumption. Identity-based consumption refers to the behavior of consumers who use their purchases to express their values and identity [11, 12]. This type of consumption is driven by the desire to create and maintain a particular image or persona, and is often influenced by cultural and social factors. Consumers may choose products or services that align with their identity or values, such as buying environment-friendly products, or purchasing luxury brands to convey their status and wealth [13].

Prior research has shown that algorithm aversion can be reduced by justifying algorithmic recommendations with explanations [2, 5]. Thereby, consumers obtain a better understanding of how a certain decision was made, which results in more positive attitudes towards an algorithm [10]. Providing explanations of algorithmic recommendations in the context of objective decision problems like medical diagnosis or financial investments seems like a rather technical problem, which has to be faced by technical disciplines such as computer science [14]. For instance, when an algorithm decides on whether a given x-ray shows infection with a disease, the algorithm could highlight the relevant parts of the image. However, in the context of identity-based consumption, the algorithm must gain an understanding of the individual motivations in regard to consumers' identity that guide their purchase decisions. In consumer research, identity is often distinguished along a social identity (i.e., what makes one similar to one's peers) and a personal identity (i.e., what makes one different from one's peers) [15]. Hence, if the provided explanation does not comply with a consumer's identity, the attempt of mitigating algorithm aversion by providing explanations can easily backfire. For example, the algorithm justifies a recommendation with the consumer's common taste, although the consumer considers themselves to be unique. Therefore, we argue that post-hoc explanations of algorithmic recommendations in identity-based consumption must be targeted towards consumers' salient identity.

In this study, we focus on purchase decisions about fashion outfits due to the prevalence of identity motives linked to fashion [16]. Consumers use fashion outfits to classify themselves and show their belonging to specific social groups [17]. Fashion is also used as an attempt to impress others [18], as well as express oneself [19].

Several online retailers rely on AI-based recommendations. The way that these recommendations are shown to the consumer varies among retailers. For instance, Asos, one of the biggest online-retailers in Europe, shows a section of fashion items denoted by"You might also like" below the current item, which emphasizes the consumers' particular taste. In contrast, the German retailer Zalando shows a list of items denoted by"Better together", which refers to aggregated preferences of others, hence a rather mainstream taste. Online purchases make up about 21% of global fashion sales with an expected annual growth of 11% until 2025 [20]. Therefore, designing effective recommender systems in the context of fashion—a particular realization of identity-based consumption—and mitigating algorithm aversion presents an important challenge for IS research and practice.

We propose a 3 $\times$ 3 between-subject experiment (activation of social identity vs. personal identity vs. no identity) $\times$ (explanations highlight unique taste vs. mainstream taste vs. no explanations) with eye-tracking. After eliciting participants general fashion preferences and activating social or personal identity, participants are shown several pairs of fashion items that are recommended from a human or an algorithm with the corresponding post-hoc explanation. Based on this, participants state their preferences towards the recommended items and their purchase intentions. Algorithm aversion is then measured as the preference for the human recommendation since the recommended pairs of outfits are always the same, but they are randomly denoted to either originate from a human or an algorithm. Furthermore, we complement the self-reported preferences by employing eye-tracking to measure preferences for fashion items based on the first fixation and total fixations [21–23].

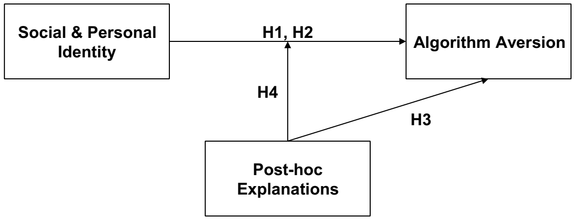
In line with self-categorization theory and theory of planned behavior, we expect that the post-hoc explanations of algorithmic recommendations need to account for consumer's salient identity in order to reduce algorithm aversion. Furthermore, we expect to find that consumers with active social or personal identity are more likely to report experience cognitive dissonance [24] or report preferences that satisfy social desirability [25–27]. Specifically, we expect that these participants are more likely to report preferences different from their true preferences. That is, the reported preferences are inconsistent with the implied preference given by the first and total number of eye fixations.

2 Prior Research and Theory

Algorithm Aversion

Algorithm aversion refers to a "biased assessment of an algorithm which manifests in negative behaviours and attitudes towards the algorithm compared to a human agent" [6, p. 5]. Algorithm aversion is often attributed to a lack of trust in the algorithm [28]. Furthermore, consumers tend to believe that algorithms are unable to account for uncertainty in decision making [9], or for a person's unique situation [7]. Moreover,

Fig. 1 Research model



humans deprive algorithms due to a lack of empathy, which makes them less likely to properly understand human behaviour and make useful recommendations [5, 10].

We analyze drivers of algorithm aversion in the context of identity-based consumption. Following our research model in Fig. 1, we develop five hypotheses that explain algorithm aversion based on the interaction of consumers' salient identity and post-hoc explanations of algorithmic recommendations. Our hypotheses are based on self-categorization theory, theory of planned behavior, social desirability, and cognitive dissonance theory.

Identity-Based Consumption

Identity as a driver of consumption has been well explored within consumer behaviour research. Consumers prefer products that are consistent with their identity [12] and choose specific products to signal that identity to others [29, 30]. Self-categorization theory (SCT) argues that in order to understand an individual's identity, one has to differentiate between their social and their personal identity [31]. Social identity describes those self-described labels that are based on being a member of certain in-groups like friends, co-workers or nationality. The personal identity refers to those self-described labels that make the individual unique (e.g., "different from my co-workers, I do not like wearing suits").

When applied to consumer behaviour, SCT is often linked to the theory of planned behaviour [TPB, 32, 33]. TPB argues that purchase intentions are particularly influenced by subjective norms and normative beliefs. Subjective norms refer to an individual's perception about the particular behavior that is widely accepted by others (e.g., "most of my friends wear shirts"), whereas normative beliefs refer to an individual's perception of social normative pressures (e.g., "professors should dress formal"). TPB further argues that these norms and beliefs are an integral part of shaping an individual's behaviour by making them consider the societal expectations and potential judgement of others [34]. Additionally, meta-analyses show that identity can be added to the TPB as a predictor of purchase intention [35].

Algorithm aversion is attributed to a lack of trust and understanding of the algorithm [28]. Specifically, algorithms lack human capabilities like empathy, which leads consumers to discount algorithmic recommendations [5, 10]. Furthermore, algorithms are categorized by humans as out-groups [36], i.e., as members of a group they are not part of themselves, which limits their influence on identity-based consumption. In particular, consumers believe that algorithms are unable to account for personal preferences or unique situations [7].

Hypothesis 1 (H1). Consumers discount algorithmic over human recommendations in identity-based consumption.

Hypothesis 2a (H2a). Activating consumers' social identity increases algorithm aversion in identity-based consumption.

Hypothesis 2b (H2b). Activating consumers' personal identity increases algorithm aversion in identity-based consumption.

Explainable AI

The field of "Explainable AI" focuses on making black-box algorithms like neural networks more accessible and understandable by humans [e.g., 14, 37, 38]. For this purpose, explainable AI research develops technical approaches that shed light onto how the decisions of AI models are made in terms of global (entire decision logic) or local explanations (for a particular input). As a result, the decisions of complex models can be approximated by less complex or even linear models. However, such approaches are not feasible in the context of online purchases as consumers cannot be assumed to spend the necessary cognitive effort in understanding technical explanations.

As a remedy, we focus on textual post-hoc explanations of algorithm recommendations. Such explanations only provide a very high-level intuition of why a recommendation was made. For instance, the explanations can highlight a mainstream taste "Others also bought" or a unique taste "You may also like." Prior studies found that algorithm aversion can be reduced by presenting explanations [2, 5]. Thereby, consumers obtain a better understanding of how a certain decision was made, which results in more positive attitudes towards an algorithm [10]. Following this line of reasoning, we expect that providing both of the aforementioned types of explanations (i.e., emphasizing mainstream or unique taste) decrease algorithm aversion. Accordingly, we propose.

Hypothesis 3 (H3). Showing explanations of algorithmic recommendations decreases algorithm aversion in identity-based consumption.

Targeted Post-hoc Explanations

We focus on targeted post-hoc explanations that either highlight the consumer's taste in relation to others (mainstream), or the consumer's particular taste (unique). Recommender systems are often based on the aggregated data of millions of transactions from other consumers [39]. For a consumer, it may hence seem reasonable to follow the mainstream and simply buy "what others bought." This behavior is consistent with the concept of social proof, which describes the tendency to copy the behaviour of others [40]. Therefore, we refer to explanations that emphasize the consumer's taste in relation to others as "mainstream explanations." Personalized recommenders, in contrast, attempt to understand the preferences and intention of each consumer individually. Instead of making recommendations based on aggregated data from others, personalized recommenders appeal to a "consumer's need for uniqueness", i.e., their "pursuit of differentness relative to others" [41, p. 1]. Accordingly, personalized recommendations provide explanations that focus on the uniqueness of the consumer, which we refer to as "uniqueness explanations."

By combining the concepts of identity-based consumption and different types of explanations, one can provide targeted mainstream explanations that appeal to a consumer's need for social proof or their need for uniqueness. Intuitively, one could argue that the explanations must be consistent with the consumers' salient identity. That is, when social identity is salient, the explanations must highlight what others bought, while, when personal identity is salient, the explanation must highlight the consumers' uniqueness.

In contrast, one could argue in the opposite direction. Consumers also have a preference for being different from others [42]. Ruvio et al. [43] have shown that when need for social approval is high, consumers seek uniqueness, in an attempt to gather approval from peers by being sufficiently different [41]. Accordingly, we argue that consumers with a salient social identity should be presented with explanations that highlight their uniqueness. To keep track of both lines of reasoning, we propose the following alternative hypotheses.

Hypothesis 4a (H4a). When Consumers' Social Identity is Activated, Showing Mainstream Explanations Decreases Algorithm Aversion, While Showing Uniqueness Explanations Increases It.

Hypothesis 4b (H4b). When Consumers' Personal Identity is Activated, Showing Uniqueness Explanations Decreases Algorithm Aversion, While Showing Mainstream Explanations Increases It.

Hypothesis 4c (H4c). When Consumers' Social Identity is Activated, Showing Uniqueness Explanations Decreases Algorithm Aversion, While Showing Mainstream Explanations Increases It.

Hypothesis 4d (H4d). When Consumers' Personal Identity is Activated, Showing Mainstream Explanations Decreases Algorithm Aversion, While Showing Uniqueness Explanations Increases It.

Cognitive Dissonance and Social Desirability

Consumers can have a general preference for human over algorithmic recommendations. However, in particular cases, the item recommended by an algorithm can still be preferred over an item recommended by a human. This can result in cognitive dissonance [24] as the consumer's general preference for human advice is in opposition to the particular preference for the item recommended by the algorithm. Consumers can then resolve the state of cognitive dissonance by spending cognitive effort and rethink their existing belief that humans generally provide better recommendations than algorithms. However, humans are generally unlikely to change existing beliefs. Instead, humans are more likely to adjust their novel preference in order to be consistent with their existing beliefs [24]. As a consequence, consumers may report a different preference from their true preference.

Furthermore, self-reports can be subject to social desirability bias [25–27]. If consumers believe that most people around them also prefer human over algorithm recommendations, they are more likely to report this preference in order to be perceived favorably by others. Therefore, consumers' reported preference can differ from their true preference due to social desirability. Given that we expect greater algorithm aversion when social or personal identity is activated, we argue that consumers with activated social or personal identity are also more likely to report preferences different from their true preferences. Accordingly, we propose.

Hypothesis 5a (H5a). Activating consumers' social identity increases the probability of consumers reporting different preferences from their true preferences in identity-based consumption.

Hypothesis 5b (H5b). Activating consumers' personal identity increases the probability of consumers reporting different preferences from their true preferences in identity-based consumption.

3 Method

We propose a 3 × 3 between-subject experiment, with a manipulation of identity activation (no identity, social identity, or personal identity is salient) and the post-hoc explanations (no explanations, mainstream, or uniqueness explanations). We focus on fashion items as underlying products due to the strong link between fashion and identity [16].

Materials

The materials are obtained from an existing dataset of fashion images.[1] We consider fashion items associated with male and female gender. We select images that only show the fashion item in isolation, without being worn by a human model. By doing so, we avoid issues that arise from participants relating to particular models, e.g., because of ethnicity. In addition, we prevent unwanted responses from participants in regard to body-image issues.

We select ten outfits for both genders for initial preference elicitation. For each outfit, we additionally select two similar outfits that subsequently present the recommendations. We perform a pre-test in an online study, so that the triples are similar to ensure that the recommendations are meaningful.

Participants

We aim to recruit university students in Bachelor and Master programs. The power analysis for a 3×3 repeated measurement between-factor ANOVA with 5 measures per participant with effect size d = 0.15, power = 0.80, and $\alpha = 0.05$ suggests a total sample size of 144, i.e., 16 participants per group.

Procedure

We first calibrate the eye-tracking device for the current participant. Subsequently, we ask the participant about whether they wear clothes associated with male or female gender. The participant is then shown ten fashion items of the respective gender in random order and asked to rate them based on a 7-point Likert scale from 1 = strong dislike to 7 = strong like. Thereby, we elicit the general fashion preferences, which later influence the recommendations. We consider scores greater or equal to 4 as positive. Each item is associated with different attributes, e.g., "street-wear." The attributes that achieved the highest scores later appear in the specific wording of the explanations.

Participants are randomly assigned to one of the three identity groups. We activate social or personal identity through the SDFF (similarities and differences from friends

[1] For instance, from kaggle. https://www.kaggle.com/datasets/paramaggarwal/fashionproduct-images-small.

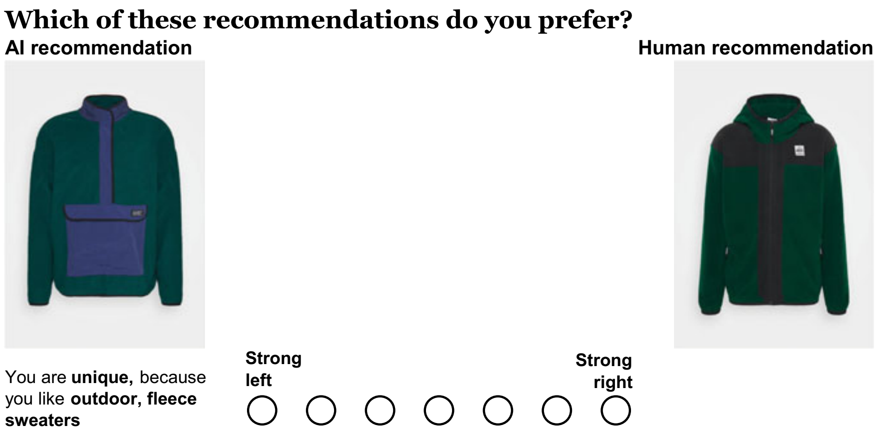


Fig. 2 Recommendation Preference Screen. The manipulated explanation is shown below the AI recommendation

and family) writing task [15]. Here, the participant is given two minutes to write down what makes them similar to (social identity) or different from (personal identity) friends and family. Participants in the control group have to write down their thoughts about different textile types. This manipulation is done after eliciting their preferences so that we do not influence their preferences through identity activation.

In the main part, participants are presented up to ten screens in random order showing recommendation pairs of one human and one algorithmic recommendation. The number of screens depends on their preferences stated at the beginning of the experiment. The screen containing the two recommendations is shown in Fig. 2. The outfit images are always presented at the same positions (i.e., left or right). However, we randomize whether the human or AI recommendation is shown on the left or right.

Which of These Recommendations Do You Prefer?

The post-hoc explanation is displayed below the image of the algorithmic recommendation. Explanations are simple "why" descriptions [44], as we cannot provide technical explanations as done by Yeomans et al. [10]. Mainstream explanations are phrased as "people similar to you like (…)," while uniqueness explanations are phrased as "you are unique, because you like (…)." The placeholders (…) are replaced by the attributes of the items that were liked during preference elicitation.

Participants are then asked to state their preference towards the items from 1 = strongly left to 7 = strongly right. Given that the actual recommendations do not actually originate from a human or an AI, we can directly measure algorithm aversion as the preference for the human recommendation. On a subsequent screen, participants are asked to state their purchase intentions for both items (irrespective of the price). The human recommendations are not framed to originate from an "expert" to avoid any biases [6]. During the experiment, we measure participants' first and total eye fixations using the Tobii Fusion eye-tracker.

Finally, participants are asked to complete a demographic survey, including their age, gender, familiarity with online purchases, disposition to trust, cultural background, program of study, and their general preference for human or algorithmic advice.

4 Expected Contributions

We expect to contribute to IS research in several ways. First, we extend the literature on algorithm aversion from high-stakes decisions in finance or medicine with an objective right or wrong [see e.g., 6] towards identity-based consumption. Due to the subjective nature of identity-based purchase decisions, developing effective recommender systems in this context is particularly challenging for IS research and practice. To determine consumers' active identity, online retailers could rely on tracking cookies to analyze consumers' online behavior prior to visiting their website. Second, we analyze how humans respond to AI explanations. However, instead of focusing on technical aspects [10, 14, 37], we study how different explanations reduce algorithm aversion in identity-based consumption. Third, we complement self-reports with eye-tracking measurements when studying algorithm aversion. Relying on self-reports only can be seen as a possible limitation when studying algorithm aversion [8]. By measuring preferences using eye-tracking, we can identify situations in which participants reported different preferences from their preferences due to cognitive dissonance or social desirability. Employing eye-tracking measurements may hence yield novel and complementary insights that self-reports alone cannot capture [26, 45].

References

1. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126.
2. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170.
3. Onkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making 22*(4), 390–409.
4. Arkes, H. R., Shaffer, V. A., & Medow, M. A. (2007). Patients derogate physicians who use a computer-assisted diagnostic aid. *Medical Decision Making, 27*(2), 189–202.
5. Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239.

6. Jussupow, E., Benbasat, I., Heinzl, A.: Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In: European conference on information systems
7. Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650.
8. Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809–825.
9. Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing, 86*(1), 91–108.
10. Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32*(4), 403–414.
11. Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology, 19*(3), 250–260.
12. Reed, A., Forehand, M., Puntoni, S., & Warlop, L. (2012). Identity-based consumer behavior. *International Journal of Research in Marketing, 29*(4):310–321.
13. Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research, 15*(2), 139–168.
14. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z.: Xai–Explainable artificial intelligence. *Science Robotics, 4*(37), eaay7120.
15. Oyserman, D., & Lee, S. (2007). Priming "culture": Culture as situated cognition
16. Davis, F. (1994). *Fashion, culture, and identity*. University of Chicago Press.
17. Loureiro, S. M. C., & de Araûjo, C. M. B. (2014). Luxury values and experience as drivers for consumers to recommend and pay more. *Journal of Retailing and Consumer Services 21*(3), 394–400.
18. Lawry, C., Choi, L., Toure, Z., & Eastlick, M. (2010). A conceptual model of ecommerce and exclusivity: Building initial trust through personal luxury values, perceived experiential value and self-concept.
19. McNeill, L., & Venter, B. (2019). Identity, self-concept and young women's engagement with collaborative, sustainable fashion consumption models. *International Journal of Consumer Studies, 43*(4), 368–378.
20. McKinsey & Company. (2023). The state of fashion 2023.
21. Ahn, J. H., Bae, Y. S., Ju, J., & Oh, W. (2018). Attention adjustment, renewal, and equilibrium seeking in online search: An eye-tracking approach. *Journal of Management Information Systems, 35*(4), 1218–1250.
22. Pfeiffer, J., Pfeiffer, T., Meißner, M., & Weiß, E. (2020). Eye-tracking-based classification of information search behavior using machine learning: Evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research, 31*(3), 675–691.
23. Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660.
24. Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
25. Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Mûller-Putz, G., Benbasat, I., Gefen, D., Kenning, P. H., Riedl, R., vom Brocke, J., & Weber, B. (2012) On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702.
26. Tams, S., Hill, K., de Guinea, A. O., Thatcher, J., & Grover, V. (2014). NeuroISalternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems, 15*(10), 723–753.
27. Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883.
28. Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting, 36*(6), 691–702.

29. Berger, J., & Heath, C.: Where consumers diverge from others: Identity signaling and product domains. Journal of Consumer Research 34, 121–134 (08 2007)
30. Cherrier, H. (2005). Using existential-phenomenological interviewing to explore meanings of consumption. *The Ethical Consumer*, 125–135.
31. Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.
32. Fielding, K. S., Terry, D. J., Masser, B. M., & Hogg, M. A. (2008). Integrating social identity theory and the theory of planned behaviour to explain decisions to engage in sustainable agricultural practices. *British Journal of Social Psychology, 47*(1), 23–48.
33. Terry, D. J., Hogg, M. A., & White, K. M. (1999). The theory of planned behaviour: Self-identity, social identity and group norms. *The British Journal of Social Psychology, 38*(3), 225–244.
34. Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.
35. Rise, J., Sheeran, P., & Sommer Hukkelberg, S. (2010). The role of self-identity in the theory of planned behavior: A meta-analysis. *Journal of Applied Social Psychology, 40*, 1085–1105.
36. Longoni, C., Cian, L., & Kyung, E. J. (2023). Algorithmic transference: People overgeneralize failures of AI in the government. *Journal of Marketing Research, 60*(1), 170–188.
37. Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The impact of explainable artificial intelligence on users' information processing. *Information Systems Research* (Forthcoming)
38. Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management, 39*(1), 53–63.
39. Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly, 31*, 137–209.
40. Cialdini, R. (1993). *Influence: Science and practice.*
41. Tian, K., Bearden, W., & Hunter, G. (2001). Consumers' need for uniqueness: Scale development and validation. *Journal of Consumer Research, 28*(1), 50–66.
42. Lang, C., & Armstrong, C. M. J. (2018). Collaborative consumption: The influence of fashion leadership, need for uniqueness, and materialism on female consumers' adoption of clothing renting and swapping. *Sustainable Production and Consumption, 13*, 37–47.
43. Ruvio, A., Shoham, A., & Brencic, M.: Consumers' need for uniqueness: Shortform scale development and cross-cultural validation. *International Marketing Review, 25*, 33–53.
44. Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 497–530.
45. Brocke, J. V., & Liang, T. P. (2014). Guidelines for neuroscience studies in information systems research. *Journal of Management Information Systems, 30*.

Take a Deep Breath and Tell Me All About It: An Experimental Study on the Effect of Breathing on Privacy Decisions

Tawfiq Alashoor, Andreas Blicher, and Rob Gleasure

Abstract This study investigates whether stress-relief breathing techniques can impact privacy decisions (i.e., the breavacy hypothesis). We asked 44 participants to complete a disclosure task consisting of 32 personal questions of low, moderate, and high sensitivity. Prior to the task, participants were assigned to a control condition, coherent breathing condition, or box breathing condition. The results reveal that participants in the box breathing condition disclosed the most personal information, followed by those in the coherent breathing condition, and the least disclosure in the control condition. The respiration data indicate that both coherent and box breathing increased the average respiration cycle duration—suggesting greater activation of the parasympathetic nervous system—with a more significant increase for box breathing than coherent breathing. Heart-rate data demonstrate that arousal is not affected by the breathing exercises. Our findings pave the way for new avenues of NeuroIS research exploring the relationship between breathing and privacy.

Keywords Breathing · Privacy decisions · Breavacy hypothesis · NeuroIS

1 Introduction

"Expression is the opposite of depression! Go for it!" exclaimed Chuck McGee III to Nestor [25, p. 147]. This occurred when Nestor "let out an uncontrolled moan" as he practiced one of the most intensive breathing techniques, Tummo (a Breathing+ technique), during his scientific journey into the realm of breath in his book, "*Breath:*

T. Alashoor (✉) · A. Blicher · R. Gleasure
Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark
e-mail: ta.digi@cbs.dk

A. Blicher
e-mail: abs.digi@cbs.dk

R. Gleasure
e-mail: rg.digi@cbs.dk

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_4

The New Science of a Lost Art" [25]. McGee argues that "self-expression" is essential for the success of Tummo, a breathing preventative therapy. Self-expression is closely related to self-disclosure. Essentially, expression necessitates communication, and communication entails revealing or disclosing one's private self through verbal and non-verbal cues [6, 12]. Self-disclosure, the extent to which people reveal themselves to others [6], is a human behavior widely studied in the privacy literature [1, 2, 15, 30].

While the link between breathing and privacy decisions may appear tenuous at first, the breathing literature [25, 37] presents a plausible connection. With approximately 22,000 breaths taken daily, humans engage in a vital process that extends beyond being a simple biochemical or physical function. Breathing profoundly influences nearly every organ, heart rate, digestion, mood, and attitude [25]. Breathing affects the parasympathetic and sympathetic nervous systems, which promote relaxation and restoration or trigger stress and emergency responses, respectively. Notably, slow breathing cycles with longer exhales induce the parasympathetic system, while fast cycles activate the sympathetic system [37]. Both systems are part of the Autonomic Nervous System (ANS), which is closely connected to various emotional responses [22, 37].

Considering the established neural links and findings from the privacy literature that highlight the significant effect of subjective and (neuro and non-neuro) objective emotional responses on privacy decisions [4, 15, 32], we suggest that stress-relief breathing techniques, such as coherent and box breathing, influence privacy decisions, such as the disclosure of personal information. We formulate and empirically examine the causal relationship between breathing and privacy decisions (i.e., the breavacy hypothesis). By examining the breavacy hypothesis, especially in light of the increasing popularity of breathing mobile apps and metaverse applications [17, 29], our research offers innovative and essential insights to the NeuroIS literature [13, 28] and the broader field of Neuro Privacy [24, 32].

The theoretical logic underlying the breavacy hypothesis can be broken down into three interconnected relationships: (1) breathing and ANS, (2) ANS and emotional responses, and (3) emotional responses and privacy decisions (see Fig. 1). First, the way we breathe directly impacts the activation of the parasympathetic and sympathetic nervous systems [20, 26]. Second, the activation of these systems within the ANS is closely related to our emotional responses [22, 37]. For example, stress-relief breathing, which activates the parasympathetic system, enhances mood and reduces levels of depression and stress [8, 27]. Finally, emotional responses have been found to significantly affect privacy decisions [15, 32], such that positive moods and emotions increase disclosure of personal information [4, 18, 21].

These relationships suggest that by manipulating breathing patterns, we can influence the ANS, emotional responses, and ultimately, privacy decisions. Specifically, we predict that stress-relief breathing will increase the level of disclosed information.

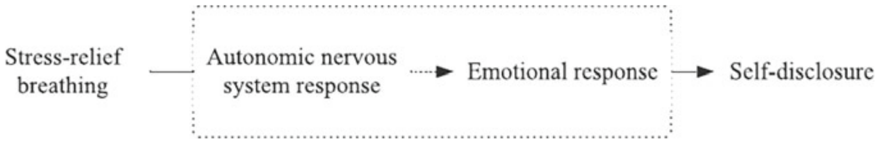


Fig. 1 Theoretical model linking breathing and self-disclosure

Hypothesis 1: Stress-Relief Breathing Techniques Positively Influence Privacy Decisions, Leading to an Increased Level of Disclosed Personal Information.

Our objective is to investigate and empirically assess the impact of breathing on privacy decisions. Utilizing objective measures of respiration and heart rate, combined with subjective measures of positive and negative affect and disclosure behavior, we observe an intricate interplay between attentional, emotional, and biological aspects of privacy decision-making. Given the exploratory nature of our research, this paper focuses on presenting the key findings from our lab experiment. By delving into the effects of breathing techniques on privacy decisions, we aim to inform individuals, organizations, and policymakers by heightening their awareness of the malleability [1, 2], context-dependence [35], or, more broadly, the general relativity of privacy decisions [5]. In the discussion section, we elaborate on the theoretical and practical implications of our findings.

2 Methods

Participants

Participants for the experiment were recruited at the authors' university ($N = 44$, $M_{\text{age}} = 23.61$, $SD = 5.25$; 33 female, 75%). The participants were given a movie ticket ($15) or course credit for their participation.

Measures

The experiment is created using Qualtrics, with iMotions used to coordinate and integrate physiological measures. The physiological measures included in the experiment were respiration (Biopac) and heart rate (Shimmer3). Respiration was measured to validate the breathing techniques. The heart rate was measured as an indicator of arousal. The following self-report measures were collected after the breathing exercise: *Disclosure Behavior* [4, experiment 2), *PANAS* [33], *Big Five Personality Factors* [19], *Disclosure Intention* [23], *Disposition to Value Privacy* [36], *Privacy*

Concerns [14], *Need for Cognition* [9], *Social Desirability* [31], and *Demographics* questions related to age, gender, education, and ethnicity.

Procedure

Participants were randomly assigned to either a control condition or one of two stress-relief breathing techniques adapted from Nestor [25]. Upon arrival at the laboratory, each participant is given verbal and written information about the experiment.[1] First, the respiration and heart-rate equipment is attached to the chest of the participant. Then, a two-minute baseline is conducted. No stimuli are presented during the baseline. After the baseline, participants in the breathing conditions take part in one of the two stress-relief breathing techniques.

Stress-Relief Treatment 1: Basic Coherent Breathing (this method reduces stress and induces a state of coherence*): "Sit up straight, relax the shoulders and belly, and exhale. Inhale to a count of 5.5; exhale to a count of 5.5. Spend 4 min breathing this way."* **Stress-Relief Treatment 2: Box (aka. Four-square) Breathing** (this method reduces stress and induces focused attention). *"Sit up straight, relax the shoulders and belly, and exhale. Inhale to a count of 4; hold 4; exhale 4; hold 4. Spend 4 min breathing this way."* During the breathing session, the participant views a video, as past research has demonstrated the efficacy of incorporating visuals into breathing training [10, 16]. The video assists the participant by presenting a dynamic circle that counts the seconds for inhalation and exhalation, resembling the motion of a Hoberman sphere, which effectively guides them through the breathing process.

Upon completion of the breathing session, the participant is asked to respond to 32 questions asking for personal information related to demographics, health, lifestyle, and ethics [4] and to respond to the other aforementioned scales. Participants in the control condition were given the chance to do one of the breathing exercises toward the end (i.e., after the disclosure and other scales) and before everyone was debriefed and thanked at the end of the experiment.

Data Analysis

The decision-making data analysis was conducted on the total number of seconds taken to disclose personal information and the total number of characters typed in response to the personal questions. The eye-tracking data analysis was conducted on the total number of fixations made during the personal questions. The decision-making data analysis and the eye-tracking data analysis was conducted on the personal questions data for all three conditions (coherent breathing condition vs.

[1] To avoid demand effects, the participants were not informed that this experiment is related to privacy decisions.

box breathing condition vs. control condition). For the eye-tracking data, an I-VT fixation filter was created. The fixation filter parameters were: 20 ms window length, 30°/second velocity threshold, 75 ms max gap length, 60 ms minimum fixation duration, 75 ms max time between fixations, and 0.5° max angle between fixations. The respiration data analysis was carried out on the average respiration cycle duration (seconds) and the heart-rate data analysis was carried out on the average R-R interval (milliseconds). The respiration data analysis and the heart-rate data analysis were conducted on the baseline and breathing session data for the two breathing conditions (coherent breathing condition vs. box breathing condition) and on the baseline data for all three conditions (coherent breathing condition vs. box breathing condition vs. control condition).[2]

3 Results

Decision-Making Data

A one-way analysis of variance (ANOVA) with group (coherent breathing condition vs. box breathing condition vs. control condition) as between-subject factor was conducted on the total number of seconds taken to respond to the personal questions. The analysis revealed a significant main effect of group, $F(2, 39) = 4.28$, $p = 0.021$, $\eta^2 = 0.18$. Post hoc independent samples *t*-tests revealed a significant difference between the box breathing condition and the control condition, $t(27) = 2.68$, $p = 0.012$, $r = 0.44$, and a significant difference between the coherent breathing condition and the control condition, $t(27) = 2.66$, $p = 0.013$, $r = 0.44$. The difference between the coherent breathing condition and the box breathing condition, $t(24) = 0.51$, $p = 0.613$, $r = 0.10$, did not reach significance. Overall, the results suggest that the number of seconds was significantly higher for the box breathing condition ($M = 508$, $SD = 185$) compared to the control condition ($M = 367$, $SD = 92$) and the number of seconds was significantly higher for the coherent breathing condition ($M = 476$, $SD = 129$) compared to the control condition ($M = 367$, $SD = 92$).

Regarding the total number of characters typed in response to the personal questions, the analysis revealed a marginally significant main effect of group, $F(2, 39) = 3.05$, $p = 0.059$, $\eta^2 = 0.14$. However, post hoc independent samples *t*-tests, which reserve degrees of freedom, revealed a significant difference between the box breathing condition and the control condition, $t(27) = 2.23$, $p = 0.037$, $r = 0.37$, and a significant difference between the coherent breathing condition and the control condition, $t(27) = 2.16$, $p = 0.040$, $r = 0.37$. The difference between the coherent

[2] Pearson correlation coefficients were computed to assess the linear relationship between respiration, number of seconds, number of characters, and number of fixations. There was a positive correlation between respiration and number of seconds, $r(33) = 0.350$, $p = 0.046$, respiration and number of characters, $r(33) = 0.316$, $p = 0.074$, and respiration and number of fixations, $r(33) = 0.394$, $p < 0.023$. These correlations help us rule out alternative explanations unrelated to breathing.

breathing condition and the box breathing condition, $t(24) = 0.79$, $p = 0.435$, $r = 0.15$, did not reach significance. Overall, the results suggest that the number of characters was significantly higher for the box breathing condition ($M = 422$, $SD = 219$) compared to the control condition ($M = 288$, $SD = 89$), and the number of seconds was significantly higher for the coherent breathing condition ($M = 368$, $SD = 109$) compared to the control condition ($M = 288$, $SD = 89$). Utilizing the standard average of 5 characters per word, participants in the box breathing, coherent breathing, and control conditions disclosed 85, 74, and 58 characters, respectively. This finding suggests that stress-relief breathing techniques increased the level of disclosed personal information, supporting the breavacy hypothesis.

Eye-Tracking Data

A one-way analysis of variance (ANOVA) with group (coherent breathing condition vs. box breathing condition vs. control condition) as between-subject factor was conducted on the total number of fixations made during the personal questions. The analysis revealed a significant main effect of group, $F(2, 36) = 6.12$, $p = 0.005$, $\eta^2 = 0.25$. Post hoc independent samples t-tests revealed a significant difference between the box breathing condition and the control condition, $t(26) = 3.16$, $p = 0.004$, $r = 0.51$. The difference between the coherent breathing condition and the control condition, $t(24) = 1.55$, $p = 0.135$, $r = 0.30$, and the difference between the box breathing condition and the coherent breathing condition, $t(22) = 1.89$, $p = 0.073$, $r = 0.37$, did not reach significance. Overall, the results suggest that the number of fixations was significantly higher for the box breathing condition ($M = 905$, $SD = 314$) compared to the control condition ($M = 599$, $SD = 193$). The number of fixations was not significantly higher for the coherent breathing condition ($M = 708$, $SD = 156$) compared to the control condition ($M = 599$, $SD = 193$).

Respiration Data

A 2×2 mixed design analysis of variance (ANOVA) with time (baseline vs. breathing exercise) as within-subject factor and group (coherent breathing condition vs. box breathing condition) as between-subject factor was conducted on the respiration data. The analysis revealed a significant main effect of time, $F(1, 18) = 148.66$, $p < 0.001$, $\eta^2 = 0.89$, a significant main effect of group, $F(1, 18) = 10.66$, $p = 0.004$, $\eta^2 = 0.37$, and a significant time $\times$ group interaction, $F(1, 18) = 18.21$, $p < 0.001$, $\eta^2 = 0.50$. Post hoc dependent samples t-tests revealed a significant difference between the box breathing condition during the breathing exercise and during the baseline, $t(9) = 15.52$, $p < 0.001$, $r = 0.97$, and a significant difference between the coherent breathing condition during the breathing exercise and during the baseline, $t(9) = 4.67$, $p < 0.001$, $r = 0.73$. Post hoc independent samples t-tests revealed a significant difference between the box breathing condition and the coherent breathing condition during the breathing exercise, $t(18) = 13.09$, $p < 0.001$, $r = 0.95$. The difference between the box breathing condition and the coherent breathing condition during the baseline, $t(18) = 0.67$, $p = 0.515$, $r = 0.15$, did not reach significance. Overall, the results suggest that the average respiration cycle was significantly longer for the box breathing condition during the breathing exercise ($M = 10.61$, $SD = 3.03$) compared to the baseline ($M = 5.94$, $SD = 0.60$), and the average respiration cycle was significantly longer for the coherent breathing condition during the breathing exercise ($M = 14.92$, $SD = 0.85$) compared to the baseline ($M = 5.21$, $SD = 1.65$). In addition, the average respiration cycle increased more for the box breathing condition ($M = 9.71$, $SD = 1.98$) than the coherent breathing condition ($M = 4.67$, $SD = 3.16$) from the baseline to the breathing exercise.

A one-way analysis of variance (ANOVA) with group (coherent breathing condition vs. box breathing condition vs. control condition) as between-subject factor was conducted on the baseline respiration data. The main effect of group, $F(2, 30) = 0.74$, $p = 0.486$, $\eta^2 = 0.05$, did not reach significance. These findings indicate that our breathing manipulations were valid, thus ensuring high internal validity for our test of the breavacy hypothesis.

Heart-Rate Data

A 2×2 mixed design analysis of variance (ANOVA) with time (baseline vs. breathing exercise) as within-subject factor and group (coherent breathing condition vs. box breathing condition) as between-subject factor was conducted on the heart-rate data. The main effect of time, $F(1, 20) = 0.79$, $p = 0.384$, $\eta^2 = 0.04$, the main effect of group, $F(1, 20) = 0.33$, $p = 0.574$, $\eta^2 = 0.02$, and the time $\times$ group interaction, $F(1, 20) = 0.12$, $p = 0.731$, $\eta^2 = 0.01$, did not reach significance.

A one-way analysis of variance (ANOVA) with group (coherent breathing condition vs. box breathing condition vs. control condition) as between-subject factor was conducted on the baseline respiration data. The main effect of group, $F(2, 34) = 0.36$, $p = 0.699$, $\eta^2 = 0.02$, did not reach significance.

4 Discussion

We demonstrate that stress-relief breathing techniques leads to an increased level of disclosed personal information and, consequently, longer response times to personal questions. Our findings suggest that when individuals are relaxed due to breathing exercises that trigger the parasympathetic nervous system [20, 25, 26, 37], they are more likely to reveal personal information compared to those in a neutral ANS state. The respiration data indicates that both coherent and box breathing increase the average respiration cycle duration, and hence activate the parasympathetic nervous system, with box breathing having a more substantial impact on self-disclosure than coherent breathing.

Our findings provide evidence for the subtle influence of breathing on privacy decisions, introducing a new NeuroIS domain with the potential to inform individuals, organizations, and policymakers about privacy decision-making. The theoretical implications of our research contribute to a deeper understanding of the interplay between physiological, emotional, and cognitive factors in privacy decisions [15]. Furthermore, our findings extend the current body of knowledge in NeuroIS [13, 28] and Neuro Privacy [24, 32] fields by identifying the role of stress-relief breathing techniques [37] in shaping privacy-related behaviors [30].

The practical implications of this research are numerous, particularly in light of increasing privacy concerns [7], heightened awareness of physiology in the workplace [34], and the growing popularity of stress management breathing apps [11, 17], including those emerging in the metaverse [29]. First, although individuals using these apps to practice stress-relief breathing techniques may experience health benefits, they should remain cognizant of the potential impact of such apps on privacy decisions. It is plausible that individuals could be more inclined to disclose personal information while interacting with other apps following a brief breathing session. Second, organizations and app developers should consider incorporating guidelines and features to promote mindful privacy practices, ensuring users are aware of the implications of their disclosure choices. For policymakers, our findings suggest the need to consider the malleability [1], context-dependence [35], or general relativity [5] of privacy decisions at the physiological level. Policies and guidelines related to personal information disclosure should take into account the subtle influences of individuals' emotional and physiological states. By promoting awareness of the effect of subtle factors on privacy decisions, policymakers can help create an environment that supports more informed and responsible privacy decision-making practices.

Our findings should be interpreted while considering several limitations. First, we used the number of characters as a proxy for measuring self-disclosure. Although this provides a nuanced measurement of disclosure behavior, previous research has shown that self-disclosure can be measured in various ways, including self-reports of intention to disclose [30] and the number of questions answered in a survey instrument like ours [4]. Due to our small sample size, we were unable to detect a significant difference in terms of the number of questions answered, as supported in Alashoor et al. [4], but we did find a significant difference in the number of characters disclosed. Our analyses addressed several alternative explanations that could potentially undermine the observed effect of breathing on this nuanced measure of disclosure behavior. Nevertheless, future research is essential to confirm our findings. Furthermore, studies investigating the impact of breathing on the content of disclosed information will offer valuable insights, as increased or decreased disclosure is not inherently unwise given the context-dependent and relative nature of privacy decisions [3, 35].

Second, we used visuals to help participants follow the breathing instructions. While such a method is encouraged by the breathing literature [10, 16], we cannot rule out the potential influence of the visuals on the baseline of disclosure behavior. While the respiration data (also see Footnote 2) provide empirical evidence for the validity of the breathing manipulation, thus rendering the visuals issue a minor limitation of the experiment, future research can refine our experimental design to further substantiate the breavacy hypothesis. Last, we did not present other analyses related to the effect of breathing on self-reported measures, such as privacy concerns, disposition to value privacy, personality factors, etc. However, we plan to embark on this project in the near future to provide a more comprehensive understanding of the breavacy hypothesis.

5 Conclusion

Irrespective of the underlying mechanism through which breathing affects decision-making, slow and deep breathing can indeed have an impact on self-expression, self-disclosure, and privacy decisions in general. Privacy decisions made in specific situations (e.g., stress) may cause individuals to express themselves less, while making those same decisions under different circumstances (e.g., relaxation) could lead to increased self-expression. This illustrates the general relativity of privacy decisions, which are influenced by a myriad of economic, psychological, and, as our study introduces, physiological factors like breathing, all of which are capable of bending the utility function of any privacy calculus depending on context and time. We hope our work serves as an impetus for future research to explore the integration of breathing and privacy, or "breavacy," considering the profound connection between the intimate nature of both concepts.

Acknowledgements This study was funded in part by a grant from the Carlsberg Foundation, grant number CF20-0254. The authors wish to thank all the study participants. The first author extends special thanks to Marie-Louise Lauritsen, whose in-depth expertise and practical experience in breathing ignited a curiosity to investigate the impact of breathing on privacy decisions. Additionally, this work was greatly inspired by James Nestor's seminal book and his invaluable contributions to the broader breathing literature, which served as a foundation for our investigation.

Declaration of Conflicting Interests The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval The study was approved by the institutional review board and was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

References

1. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509–514.
2. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology, 30*(4), 736–758.

3. Alashoor, T. (2024). It is seriously time for privacy education, training, and awareness (PETA) programs. In B. James & M. Robles (Eds.), *Effective methods for teaching business related topics during and post crisis*. National Business Education Association.
4. Alashoor, T., Keil, M., Jeff Smith, H., & McConnell, A. R. (2023). Too tired and in too good of a mood to worry about privacy: Explaining the privacy paradox through the lens of effort level in information processing. *Information Systems Research*, *34*(4), 1415–1436.
5. Alashoor, T., Aldawood, H., Alsulayyim, B., Almutairi, S., & Alotaibi, E. R. (2023). An online randomized field experiment on the importance of privacy education, training, and awareness (PETA). In *Proceedings of the 3rd International Conference on Computing and Information Technology IEEE*. Tabuk, Saudi Arabia.
6. Altman, I. (1975). *The environment and social behavior: Privacy, personal space, territory, and crowding*. Brooks/Cole.
7. Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019). *Americans and privacy: Concerned, confused and feeling lack of control over their personal information*. The Pew Research Center. Retrieved February 25, 2020, from https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/
8. Busch, V., Magerl, W., Kern, U., Haas, J., Hajak, G., & Eichhammer, P. (2012). The effect of deep and slow breathing on pain perception, autonomic activity, and mood processing—An experimental study. *Pain Medicine, 13*(2), 215–228.
9. Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131.
10. Chittaro, L., & Sioni, R. (2014). Evaluating mobile apps for breathing training: The effectiveness of visualization. *Computers in Human Behavior, 40*, 56–63.
11. Coulon, S. M., Monroe, C. M., & West, D. S. (2016). A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. *American Journal of Preventive Medicine, 51*(1), 95–105.
12. Debatin, B., Lovejoy, J. P., Horn, A. K., & Hughes, B. N. (2009). Facebook an online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer-Mediated Communication, 15*(1), 83–108.
13. Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research, 22*(4), 687–702.
14. Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61–80.
15. Dinev, T., McConnell, A. R., & Smith, H. J. (2015). Research commentary—Informing privacy research through information systems, psychology, and behavioral economics: Thinking outside the "APCO" box. *Information Systems Research, 26*(4), 639–655.
16. Faust-Christmann, C. A., Taetz, B., Zolynski, G., Zimmermann, T., & Bleser, G. (2019). A biofeedback app to instruct abdominal breathing (breathing-mentor): Pilot experiment. *JMIR mHealth and uHealth, 7*(9), e13703.
17. Fleming, N. (2020). Are breathing techniques good for your health? *The Guardian*. Retrieved April 5, 2023, from https://www.theguardian.com/lifeandstyle/2020/jul/12/are-breathing-techniques-good-for-your-health
18. Forgas, J. P. (2011). Affective influences on self-disclosure: Mood effects on the intimacy and reciprocity of disclosing personal information. *Journal of Personality and Social Psychology, 100*(3), 449–461.
19. Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528.
20. Jerath, R., Edry, J. W., Barnes, V. A., & Jerath, V. (2006). Physiology of long pranayamic breathing: Neural respiratory elements may provide a mechanism that explains how slow deep breathing shifts the autonomic nervous system. *Medical Hypotheses, 67*(3), 566–571.
21. Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: The effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal, 25*(6), 607–635.

22. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421.
23. Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research, 15*(4), 336–355.
24. Meshi, D., Tamir, D. I., & Heekeren, H. R. (2015). The emerging neuroscience of social media. *Trends in Cognitive Sciences, 19*(12), 771–782.
25. Nestor, J. (2020). *Breath: The new science of a lost art*. Penguin.
26. Pal, G. K., Velkumary, S., & Madanmohan. (2004). Effect of short-term practice of breathing exercises on autonomic functions in normal human volunteers. *Indian Journal of Medical Research*, *120*(2), 115–121.
27. Perciavalle, V., Blandini, M., Fecarotta, P., Buscemi, A., Di Corrado, D., Bertolo, L., Fichera, F., & Coco, M. (2017). The role of deep breathing on stress. *Neurological Sciences, 38*(3), 451–458.
28. Riedl, R., Fischer, T., Léger, P. M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 51*(3), 13–54.
29. Rumage, J. (2022). Mental health in the metaverse: Mind-Easy opens 'clinic' in Decentraland. Retrieved April 5, 2023, from https://builtin.com/media-gaming/mental-health-metaverse
30. Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1015.
31. Stöber, J. (2001). The social desirability scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222–232.
32. Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, 109*(21), 8038–8043.
33. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.
34. Whelan, E., McDuff, D., Gleasure, R., & Vom Brocke, J. (2018). How emotion-sensing technology can reshape the workplace. *MIT Sloan Management Review, 59*(3), 7–10.
35. Xu, H., & Zhang, N. (2022). From contextualizing to context theorizing: Assessing context effects in privacy research. *Management Science, 68*(10), 7383–7401.
36. Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information privacy concerns: Linking individual perceptions with institutional privacy assurances. *Journal of the Association for Information Systems, 12*(12), 798–824.
37. Zaccaro, A., Piarulli, A., Laurino, M., Garbella, E., Menicucci, D., Neri, B., & Gemignani, A. (2018). How breath-control can change your life: A systematic review on psycho-physiological correlates of slow breathing. *Frontiers in Human Neuroscience, 12*, 353.

AI-Assisted Hate Speech Moderation—How Information on AI-Based Classification Affects the Human Brain-In-The-Loop

Nadine R. Gier-Reinartz, Vita E. M. Zimmermann-Janssen, and Peter Kenning

Abstract Every day, social media content moderators must decide within seconds hundreds of times whether or not user-generated content constitutes hate speech. Although IS research is making continual progress in automatically detecting potential hate speech content through AI-assisted processing, the final decision still resides in the human-in-the-loop. To support the content moderators, the results of AI-based classifications are regularly displayed during the decision-making process—but is this advisable? To approach an answer, the neural and behavioral effects of two opposing AI-based classifications are tested against each other. The results from a fNIRS experiment show that opposing AI-based classifications leads to different cortical activation patterns, which in turn depend on the individual's importance of hate speech prevention. Moreover, this exploratory study indicates that AI-based classifications may also induce a "cortical relief" seemingly cause behavioral effects that at least cast doubt on the validity and desirability of the AI-assisted human decision.

Keywords fNIRS · Hate speech detection · Content moderation · Human-in-the-Loop · Artificial intelligence · Decision support systems

---

N. R. Gier-Reinartz (✉) · V. E. M. Zimmermann-Janssen · P. Kenning
Chair of Business Administration, esp. Marketing, Heinrich Heine University, Düsseldorf, Germany
e-mail: nadine.gier@hhu.de

V. E. M. Zimmermann-Janssen
e-mail: vita.zimmermann@hhu.de

P. Kenning
e-mail: peter.kenning@hhu.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_5

1 Introduction

Given the tremendous amount of user-generated content, the initial detection and subsequent moderation of potential online hate speech[1] (HS) pose an ongoing challenge for social media platforms [1, 2]. Every day, social media content moderators must decide whether or not user-generated content constitutes HS within a few seconds—many hundreds of times [3, 4]. Little is known about the actual process of content moderation in social media. However, based on interviews with former content moderators, a sense of the process can be obtained [3]. In the process, moderators receive randomly assigned reports (tickets) consisting of diverse texts, images or videos. It is unpredictable what will appear next on the screen, which makes the situation stressful, as it is impossible to prepare psychologically. Due to strict productivity metrics, moderators have only a few seconds to decide or rather react according to company guidelines. This places immense physical and psychological demands on content moderators, which not only bear the responsibility for deciding whether the content is (not) HS and thus ultimately (not) banned from the platform [1]. It is also them to make these difficult and context sensitive decisions at the fuzzy boundary of free speech and hate speech [3, 5].

Instead of exploiting contract labors [3] or workforces in mainly poorer countries for this process at the expense of their mental health, as seems to have been the case in the past, social media companies now seem to be turning more to AI solutions [6]. In this regard, valuable progress has been made in IS research in the detection of HS content. Besides democratic crowd-based approaches [7], a multitude of efficiency-increasing semi- and fully automated AI solutions for HS detection are proposed in IS research [8, 9]. However, when it comes to the subsequent moderation of this content, alike research on mechanisms is largely missing for the moderation phase [1]. It seems that there is, besides legal restrictions (e.g, Art. 22 General Data Protection Regulation (GDPR)), to date no sufficient workaround to completely forgo the human-in-the-loop as the final decision-making authority in this moderation phase [10]. Thus, content moderators still decide on the final treatment of content. Yet it is precisely this moderation phase that demands increased cognitive effort and psychological strain [4]. In this process, the display of AI-based classifications (AIC) is suggested to help content moderators with their decision [11]. Moreover, new IS approaches propose that the final decisions of the content moderator can subsequently serve as new data samples for training AI classification [12]. However, is it advisable to display AIC to the content moderator during the decision-making process if it is still unclear if and how the display of AIC affects the human-in-the-loop? In this regard, at least two questions remain open: (1) To what extent does the display of AIC affect the moderator's decision, and (2) does the displayed AIC bias the decision-making process by involving varying cortical neural processes?

To address these two questions, the remainder of the paper will first outline which processess could possibly be relevant in this regard and how they can be measured

[1] Hate speech can be defined as communication "that attacks or uses pejorative or discriminatory language with reference to a person or a group […] based on identity factors" (p.10, [37]).

using neurophysiological methods. Then the method is described and it is explained which participants were recruited and how the experimental task and the corresponding stimuli were constructed. After describing the (pre-)analysis procedures, the results are presented. Finally, these are discussed against the background of possible theoretical and practical implications.

2 Theoretical and Methodological Considerations

To observe the behavioral and neural responses to the displayed AIC during decision-making in context of hate speech an explorative, ethically approved, event-related functional near-infrared spectroscopy (fNIRS) experiment was executed. Since it is thought that the displayed AIC can reduce cognitive effort for content moderators, an exploratory neural measure of related constructs such as mental workload or cortical relief seem to be a good starting point. Albeit electroencephalography (EEG) has been commonly used to quantify mental workload in the past given its high temporal resolution that allows for on-time measurements [13, 14], fNIRS has recently gained attention as complementary method [13, 15, 16]. This is because fNIRS can quantify multiple hemodynamic and metabolic parameters and identify the cortical origin of neural activity, allowing to identify neural cortical pattern comparable to functional magnetic resonance imaging (fMRI) results [17].

A characteristic neural pattern that may be of interest in the context of content moderation is the cortical relief effect [18–20]. This effect is a specific neural mechanism in human decision making. It is defined by increased activation in areas of the medial prefrontal cortex (mPFC) and concomitant decreased activation in areas of the lateral PFC (lPFC). Thereby, the increased mPFC activation is assumed to be associated with increased subjective appraisal and simultaneously less cognitively engaging and less complex neural processing as suggested by lPFC deactivation [21, 22]. The activation pattern was originally identified with fMRI during a binary decision task when choosing between two branded products [20]. In choice sets where the most preferred brand is available as a choice compared to sets where there are two non-favored product brands to choose from this characteristic neural pattern was found. The interpretation is that a strongly preferred brand simplifies the choice decision and cortically "relieves" the decision-making process [19, 20]. The cortical relief effect has been replicated with fNIRS, also being used to correlate with real-world data [18, 19]. Since this study aims as a first step to proof whether the display of AIC has the intended relieving effect on the decision-making process, it seems sensible to first explore a validated effect with a district neuronal pattern. For this purpose, it is necessary to be able to differentiate cortical areas. At the same time, it should be possible to easily transfer the results to real, on-time applications for later studies. Thus, the application of fNIRS seems to be a promising first approach in this direction.

Participants

Participants who were at least 18 years old were recruited from the institutional participant pool. They were only included if they were neither pregnant nor breastfeeding to (1) avoid possible emotional distress for these vulnerable groups and (2) circumvent additional variance between participants due to hormonal and neurochemical imbalances [40]. For latter reason, also participants who were taking (neuro-)psychoactive medication or did suffer from severe mental or neurological disorders were excluded. Furthermore, participants should at least have one active social media profile. Those with a presumably extreme social desirability bias, indicated by most extreme scores on the scale of socially desirable response behavior [23], were excluded for analysis ($n = 2$). This resulted in a final sample of $n = 28$ ($M_{\text{age}} = 27.07(9.63)$, 64.3% female).

Material

For stimuli development, potential HS comments had to be created first, fulfilling the three quality criteria of *credibility*, *offensiveness*, and *comprehensibility*, that are considered particularly influential in the applied study design. As an initial sample, 165 comments were pretested, created from commonly used hate expressions [24] and targets of different hate categories (from https://hatebase.org). In the pretest ($N = 173$, $M_{\text{age}} = 42.71(12.24)$, 59.4% male) 30 randomly chosen comments were rated per participants on the three criteria on 10-point scales. *Credibility* was assessed using an item on 'believability' [25, 26], including only comments with an average score of eight or higher. The focal criterium of *offensiveness* was assessed using three items that provided information on how 'insulting', 'discriminatory', and 'offensive' each comment was perceived to be. The selected offensiveness scores should be within the interval of 7.0–8.5 points in order to select HS comments in a way that will cause some variance in the perception of the offensiveness. *Comprehensibility* was ensured by only considering comments where at most one participant marked an incomprehensible word within the comment. Finally, 60 comments were identified that were used with an additionally displayed AIC below the comment as stimuli in the experimental task. The AIC was an additional information, fictitiously classifying the comment as problematic (pr) or non-problematic (npr) ("AI analysis: comment [not] problematic"). Participants had to decide whether they want to flag the potential HS comment as problematic. Twelve comments were randomly assigned to each AIC condition, resulting in 24 trials for the AIC conditions which were displayed in randomized order.

Neural data was acquired with the non-invasive mobile neuroimaging method of fNIRS fitted in a headband. This fNIRS was attached to the participant's forehead using the craniometric point of the nasal bone as an application reference to ensure equal device positioning across participants. The continuous-wave fNIRSport-System (NIRx Medical Technologies, Berlin) was used, collecting the

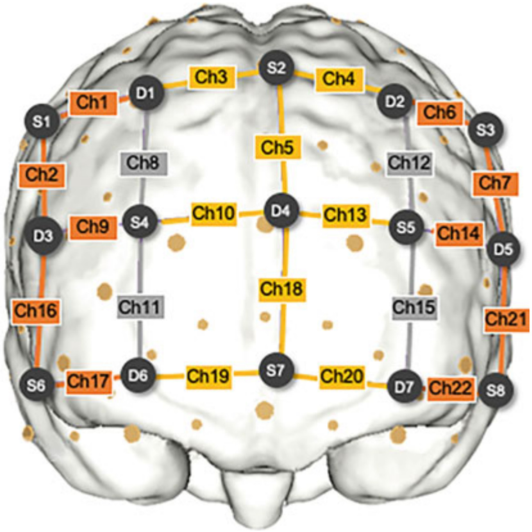


Fig. 1 Schematic visualization of the fNIRS montage and channel localization. The coloring of the channels shows the classification of the associated brain regions. Orange = lateral PFC, yellow = medial PFC and grey are unclassified channels. D = light detector; S = light source; Ch = channel

data with the NIRS-Star software package (v14.2). By penetrating the human tissue with near-infrared light (wavelengths of 760 and 850 nm), de-/oxygenated blood flow in subjacent cortical brain areas was measured via levels of hemoglobin concentration (HbO and HbR; [27]). The data were sampled at a 7.81 Hz, using 22 channels (resulting out of 8 light source emitters and 7 long-distance light detectors; average distance 30 mm, Fig. 1) covering the medial (3–5, 10, 13, 18–20) and lateral prefrontal cortex (1, 2, 6, 7, 9, 14, 16, 17, 21, 22). For four channels (8, 11, 12, 15) the grouping is too vague.

Experimental Procedure

After participants were informed about the study and the fNIRS device in writing and verbally, informed consent was obtained according to the Declaration of Helsinki [41]. After calibrating the fNIRS signal strength and shielding the device from external light with a darkening hood, the participants could self-start the task on a computer screen.

First, the comment was displayed without additional information (comment-only [CO]; 7 s). After a randomized jitter (interstimulus interval [ISI]; 1–5 s), the additional AIC was displayed with the comment (comment-AIC [CAIC]; 7 s). After another ISI (1–5 s), participants had to decide whether they want to flag the comment as HS or not, whereby first only the question was shown (2 s) before the answers (flag comment/do nothing) could be entered (2 s). For flagged comments, two additional questions had to be answered (no time restriction) regarding the reason for flagging (racism, violence, discrimination, incitement, insult, other) and the discriminated

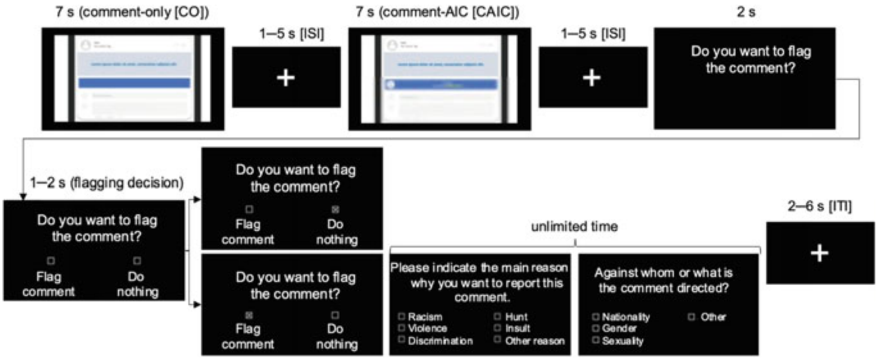


Fig. 2 Schematic trial of the experimental task

group (nationality, gender, sexuality, other). The next trial started after a jittered intertrial interval (ITI; 2–6 s). A schematic display of a trial is provided in Fig. 2.

After having completed the experimental task, the fNIRS device was removed and the participants were handed a questionnaire to query sociodemographic data and potential control variables such as *warm glow* [28], *anticipated guilt* [29], and *importance of HS prevention*, for which the harm of hate speech scale served as proxy [30]. Participation was compensated (€30) and took about one hour on average.

Preprocessing and Data Analysis

The NIRS AnalyzIR toolbox was used for preprocessing in MATLAB [31]. First, the raw fNIRS signal was sampled to 4 Hz [32]. Thereafter, data were smoothed and artifacts (e.g., heart rate or drifts in the optical signal) were removed. A baseline correction and motion correction was done, using a spatial principal component filter for the latter one [33]. Then, the optical density was calculated [32], before the signal was converted with the modified Beer-Lambert law (partial pathlength factor of 0.1) into hemoglobin values [34]. For each participant, a general linear model (GLM) was calculated, modeling the experimental periods (CO, CAIC, decision) as regressors with a duration according to their experimental phase.

After GLM estimation with the AR-IRLS algorithm [35], each time course was convolved by the canonical hemodynamic response function. A mixed-effects model was calculated for group analysis, using AIC conditions as fixed effects and subjects as random effects.

The contrast of interest was calculated between the AIC conditions (npr < pr) each being controlled for the activation patterns in the comment-only period (CO < CAIC) to see whether different AIC results in different processing. The subsequent results will be reported if significant (family-wise error corrected threshold $q < 0.05$)

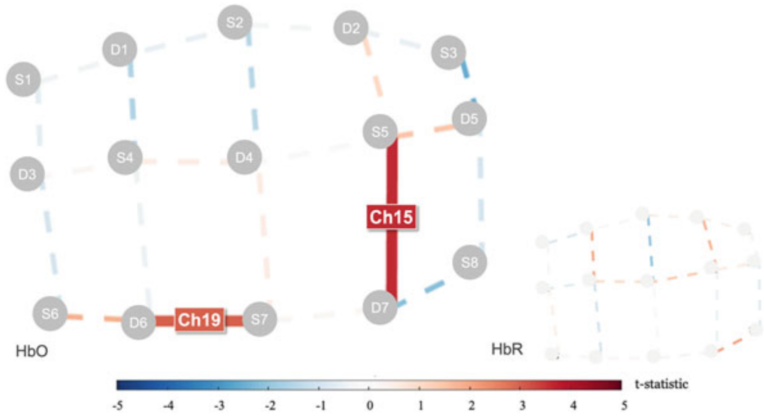


Fig. 3 Significant channels of the contrast comparing the AIC conditions ([CO-npr<CAIC-npr]<[CO-pr<CAIC-pr]). The color coding corresponds to the t-values indicated by the color bar. Significant activation in uncategorized channels: Ch15: HbO $t(27) = 3.658$, $q = 0.01$; HbR $t(27) = -0.422$, $q = 0.857$. D = light detector; S = light source; Ch = channel; HbO = oxygenated hemoglobin; HbR = deoxygenated hemoglobin

and robustness will be checked (HbO and HbR suggest the same neural activation within one channel, [36]).

3 Results

Contrasting the AIC conditions (npr < pr) identified a significant increase in the medial PFC (Ch19: HbO $t(27) = 3.0197$, $q = 0.039$; HbR $t(27) = 0.177$, $q = 0.94$; Fig. 3).

Similarly, the behavioral results suggest different responses to AIC conditions. The applied repeated measures (RM-)ANOVA[2] reveals a significant main effect of AIC on the proportion of flagged comments (propFlag), $F(1, 27) = 12.993$, $p \leq 0.001$, $\eta_p^2 = 0.325$, indicating an AIC compliant flagging behavior with 75% of the comments being flagged when classified as problematic by the AI compared to non-problematic (62%). This effect stays positive and significant when controlling for sociodemographic data, $F(1, 23) = 20.467$, $p \leq 0.001$, $\eta_p^2 = 0.471$, and additional latent variables, with one exception: The inclusion of *importance of HS prevention* (HSP) results in a marginally nonsignificant main effect of AIC.

[2] The assumptions of RM-ANOVA were tested. Sphericity is given since there are only two conditions. The residuals were not normally distributed per condition, as indicated by significant Shapiro–Wilk tests. Though, not only is RM-ANOVA generally considered robust to violations of normality in terms of type I and type II error, as shown by systematic Monte Carlo simulation studies of F-test robustness considering a wide variety of distributions, but also specifically the applied sample size of $n = 28$ exceeding the threshold commonly suggested for robustness [38, 39].

Since HSP has already been shown to be a potential moderator of the behavioral response in flagging decision contexts [7], a post-hoc analysis was performed. The split-plot RM-ANOVA[3] with HSP as a 2-level between-subject factor (median-split HSP groups) reveals a significant interaction of AIC × HSP, $F(1, 26) = 15.925$, $p = 0.002$, $\eta_p{}^2 = 0.305$, being robust under control for sociodemographic data, $F(1, 22) = 6.514$, $p = 0.018$, $\eta_p{}^2 = 0.228$. While this already suggests that AIC affects HSP groups differently, follow-up groupwise inspection of RM-ANOVA confirms that those with low levels of HSP (HSP_{low}) are particularly impacted by AIC and behave compliant to AIC, $F(1, 14) = 22.770$, $p \leq 0.001$, $\eta_p{}^2 = 0.619$. Those with high levels of HSP (HSP_{high}) show no behavioral response in turn, $F(1, 12) = 0.263$, $p = 0.618$, $\eta_p{}^2 = 0.021$.

Inspired by this moderation effect found on behavioral level, the interaction contrast of the AIC main effect between the HSP groups was re-analyzed on neural level.

Surprisingly, the interaction contrast (Fig. 4A) identifies stronger significant and robust deactivation in the right lateral PFC (Ch2: HbO $t(27) = -2.33$, $q = 0.13$; HbR $t(27) = 3.333$, $q = 0.019$; Ch17: HbO $t(27) = -2.169$, $q = 0.157$; HbR $t(27) = 3.145$, $q = 0.029$) and increased significant activation in the medial PFC (Ch19: HbO $t(27) = 2.996$, $q = 0.04$; HbR $t(27) = 0.371$, $q = 0.88$) for HSP_{high} compared to HSP_{low}, suggesting that a cortical relief effect specifically arises for those with high levels of HSP. Simple effect contrasts confirm this finding. The significant increased mPFC (Ch19: HbO $t(27) = 3.079$, $q = 0.034$; HbR $t(27) = 0.361$, $q = 0.883$) was mainly found for HSP_{high} (Fig. 4B2) while for HSP_{low} no significant activation could be identified (Fig. 4B1).

4 Discussion

Final decisions in content moderation place a tremendous psychological burden on content moderators. It is assumed that the display of AIC supports the moderators in this decision, wherein this work-in-progress study aimed to show how much the decision and also the decision-making process are affected by such information on AI classification. Specifically, two research questions were addressed: (1) To what extent does the display of AIC affect the moderator's decision, and (2) does the displayed AIC bias the decision-making process by involving varying cortical neural processes? To answer the first question, this study shows that the decision to flag a potentially problematic comment can be influenced by information on AIC and differs depending on the individual importance of HSP. In particular, the display of

[3] The assumptions of split-plot RM-ANOVA were tested. For sphericity and normal distribution of residuals per condition please see Footnote 2. Box's M of 4.559, $F(3, 484{,}765.792) = 1.392$, $p = .243$, indicates equality of the covariance matrices of propFlag across the cells formed by HSP groups. Homogeneous error variances of propFlag could be assumed between HSP groups as assessed using Levene's test per condition, AIC_{pr}: $F(1, 26) = 0.747$, $p = .395$, and AIC_{npr}: $F(1, 26) = 4.057$, $p = .054$.

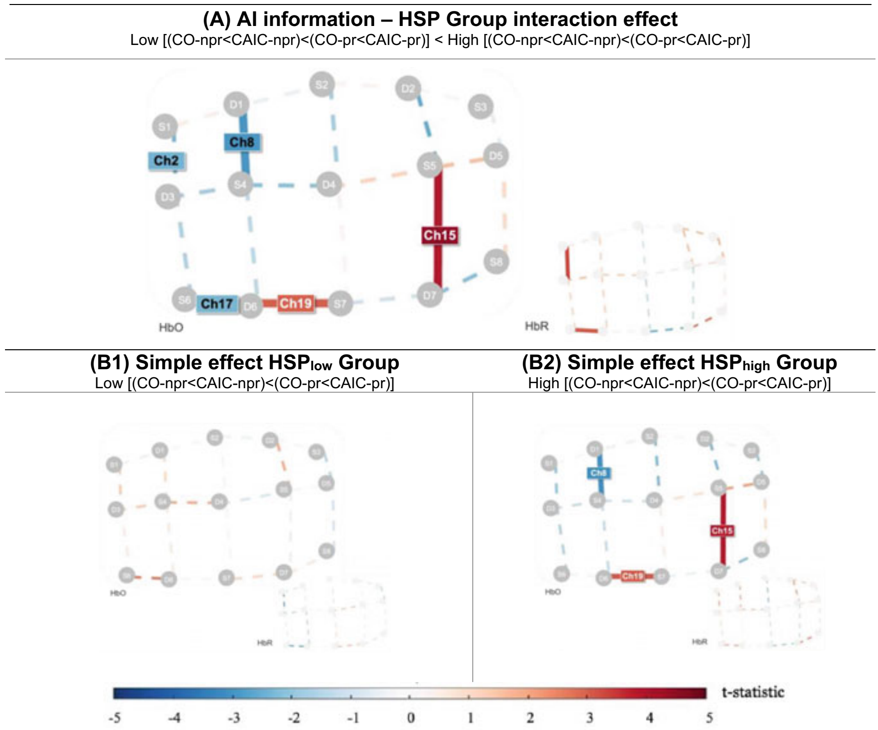


Fig. 4 Significant channels of the contrast displaying (A) the interaction of AIC with HSP group and (B1/B2) the AIC contrast per group. The color coding corresponds to the t-values indicated by the color bar. Significant activation in uncategorized channels: (A) Ch8: HbO $t(27) = -2.926$, $q = 0.046$; HbR $t(27) = 1.058$, $q = 0.56$; Ch15: HbO $t(27) = 3.959$, $q = 0.005$; HbR $t(27) = -1.15$, $q = 0.511$; (B2) Ch8: HbO $t(27) = -3.092$, $q = 0.033$, HbR $t(27) = 1.98$, $q = 0.206$; Ch15: HbO $t(27) = 3.981$, $q = 0.005$; HbR $t(27) = -1.178$, $q = 0.501$. D = light detector; S = light source; Ch = channel; HbO = oxygenated hemoglobin; HbR = deoxygenated hemoglobin

AIC seems to primarily affect the decisions of individuals that place less importance on HSP, while individuals with high importance of HSP seem to be behaviorally unaffected by the displayed AIC. However, with regard to neural responses, it is the exact opposite and specifically this latter group of individuals with high importance of HSP seems to profit neurally from the display of AIC. However, this cortical relief effect specifically in the mPFC cannot be found for those with comparably low importance placed on HSP.

These results indicate that, on the one hand, the decision-making process of individuals who attach high importance to HSP is impacted by the displayed AIC but, on the other hand, their actual decision seems to be unaffected by any displayed AIC. These decision-makers seem to benefit from AIC in the desired way by being more cortically relieved during the decision-making process compared to the individuals who attach low importance to HSP, especially when the displayed AIC matches their general repudiation of hate speech. However, the actual decision of individuals with high importance to HSP does not seem to be influenced by the display of the AIC, as their flagging behavior is similar in both AIC conditions. In contrast, the individuals who attach less importance to HSP rather seem to "blindly" follow the suggestions of the displayed AIC and do not experience any greater relief from either piece of information. Though, these individuals might be the ones who actually reflect a real content moderator to a large extent, as the enormous workload under time pressure and the constant confrontation with HS might diminish the sensitivity to HS.

However, if their "blind" decisions are in turn used as feedback to retrain the AI as suggested by new approaches [12], the question arises whether display of AIC ultimately counteracts the human-in-the-loop principle. The principle is actually intended to objectify the process by having the human moderator as the AI's control mechanism and thereby providing new qualified, human-controlled input to the system. This input is then used as a new data sample for the training set, on which the AI can improve its classification. However, if the display of AIC does not support human decision-making but rather overrides it, in that the human "blindly" follows the AIC, this could lead to biased training sets that again only replicate the AIC. If the AI now learns from this biased set, the bias becomes larger and larger, making the human-in-the-loop obsolete.

Future research should pay more attention to this possibly diametrical effect, as it has important implications for research, AI development and content moderation in practice. Therefore, additional analyses of the data, e.g., correlation analyses combining neural and behavioral data, are recommended to provide additional insights into the effect. Furthermore, replication of these exploratory effects is needed to verify the robustness of the results, where, for example, the type of AIC information and explanations of the AIC process could be manipulated. Future research could examine the effect by mimicking the actual process of content moderation on a larger scale, e.g., with real content moderator and an actual deletion decision, additionally increasing cognitive workload in advance. Furthermore, the effect could be replicated with higher temporal resolution measurements (e.g., EEG) to potentially develop an application to measure this bias within the decision-making process real-time.

Acknowledgements The funding of this study was supported by the Jürgen Manchot Stiftung.

References

1. Gongane, V. U., Munot, M. V., & Anuse, A. D. (2022). *Detection and moderation of detrimental content on social media platforms: current status and future directions*. Springer.
2. Wilson, R. A., & Land, M. K. (2020). Hate speech on social media: Content moderation in context. *Connecticut Law Review, 52*, 1029–1242.
3. Gültekin-Punsmann, B. (2018). *Three months in hell: What I learned from three months of content moderation for Facebook in Berlin*, https://sz-magazin.sueddeutsche.de/internet/three-months-in-hell-84381

4. Newton, C. (2019). *The trauma floor: The secret lives of Facebook moderators in America.* https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona
5. Kapil, P., & Ekbal, A. (2020) A deep neural network based multi-task learning approach to hate speech detection. *Knowledge-Based System, 210*, Article 106458.
6. Criddle, C., & Murgia, M. (2023). *Meta's east African content moderation hub shuts down.* https://www.ft.com/content/6c2ddd5b-1593-45c0-95b5-b35f3711f288
7. Zimmermann-Janssen, V. E. M., & Gier, N. R. (2022). Chances and limits of community-based hate speech detection—Results from a combined behavioral-neuroIS study. In *ICIS 2022 Proceedings* (p. 14).
8. MacAvaney, S., Yao, H. R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. *PLoS One, 14*, Article e0221152.
9. Meske, C., & Bunde, E. (2022) Design principles for user interfaces in AI-based decision support systems: The case of explainable hate speech detection. *Information Systems Frontiers, 25(*2), 743–773.
10. Commission, E. (2018). Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online. *Official Journal of the European Union, 63*, 50–61.
11. Darbinyan, R. (2022). *The growing role of AI in content moderation.* https://www.forbes.com/sites/forbestechcouncil/2022/06/14/the-growing-role-of-ai-in-content-moderation/
12. Kotarcic, A., Hangartner, D., Gilardi, F., Kurer, S., & Donnay, K. (2022) Human-in-the-Loop hate speech classification in a multilingual context. https://doi.org/10.48550/arXiv.2212.02108
13. Peck, E. M., Afergan, D., & Yuksel, B. F., Lalooses, F., & Jacob, R. J. K. (2014). Using fNIRS to measure mental workload in the real world. *Advances in Physiological Computing*, 117–139.
14. So, W. K. Y., Wong, S. W. H., Mak, J. N., & Chan, R. H. M. (2017). An evaluation of mental workload with frontal EEG. *PLoS One, 12*, e0174949.
15. Causse, M., Chua, Z., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Science and Reports, 7*, 1–15.
16. Unni, A., Ihme, K., Surm, H., Weber, L., Ludtke, A., Nicklas, D., Jipp, M., & Rieger, J. W.: Brain activity measured with fNIRS for the prediction of cognitive workload. In *6th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2015—Proceedings* (pp. 349–354).
17. Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage, 85*, 6–27.
18. Gier, N. R., Strelow, E., & Krampe, C. (2020). Measuring dlPFC signals to predict the success of merchandising elements at the point-of-sale—A fNIRS approach. *Frontiers in Neuroscience, 14*.
19. Krampe, C., Gier, N. R., & Kenning, P. (2018). The application of mobile fNIRS in marketing research—Detecting the 'first-choice-brand' effect. *Frontiers in Neuroscience, 12*, Article 433.
20. Deppe, M., Schwindt, W., Kugel, H., Plaßmann, H., & Kenning, P. (2005). Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *Journal of Neuroimaging, 15*, 171–182.
21. Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427.
22. Dixon, M. L., & Christoff, K. (2014). The lateral prefrontal cortex and complex value-based learning and decision making. *Neuroscience and Biobehavioral Reviews, 45*, 9–18.
23. Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2012). Eine Kurzskala zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens [A short scale to assess the gamma factor of socially desirable response rehavior]. *GESIS-Working Paper, 25*, 1–27.

24. Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. In *Proceedings of Tenth International AAAI Conference Web and Social Media (ICWSM 2016)* (pp. 687–690).
25. Sungur, H., Hartmann, T., & van Koningsbruggen, G. M. (2016). Abstract mindsets increase believability of spatially distant online messages. *Frontiers in Psychology, 7*, Article 1056.
26. Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly, 93*, 59–79.
27. Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage, 63*, 921–935.
28. Ferguson, E., & Flynn, N. (2016). Moral relativism as a disconnect between behavioural and experienced warm glow. *Journal of Economic Psychology, 56*, 163–175.
29. Theotokis, A., & Manganari, E. (2015). The impact of choice architecture on sustainable consumer behavior: The role of guilt. *Journal of Business Ethics, 131*, 423–437.
30. Cowan, G., Resendez, M., Marshall, E., & Quist, R. (2002). Hate speech and constitutional protection: Priming values of equality and freedom. *Journal of Social Issues, 58*, 247–263.
31. Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS brain AnalyzIR toolbox. *Algorithms, 11*, Article 73.
32. Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics, 3*, Article 010401.
33. Zhang, Y., Brooks, D. H., Franceschini, M. A., & Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *Journal of Biomedical Optics, 10*, Article 011014.
34. Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer-Lambert law revisited. *Physics in Medicine & Biology, 51*, 91–98.
35. Barker, J. W., Aarabi, A., & Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical Optics Express, 4*, 1366–1379.
36. Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods, 22*, 46–68.
37. United Nations. (2020). United Nations strategy and plan of action on hate speech: Detailed guidance on implementation for United Nations field presences. United Nations Report 1–52.
38. Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology, 6*, 147–151.
39. Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*, 552–557.
40. Moya, J., Phillips, L., Sanford, J., Wooton, M., Gregg, A., & Schuda, L. (2014). A review of physiological and behavioural changes during pregnancy and lactation: potential exposure factors and data gaps. *Journal of Exposure Science & Environmental Epidemiology, 24*(5), 449–458.
41. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013 Nov 27;310(20):2191-4. https://doi.org/10.1001/jama.2013.281053. PMID: 24141714.

Do Generative AI Tools Foster Positive Experiences in Knowledge Work? A NeuroIS Research Proposal

Michael Thomas Knierim and Lorenzo del Puppo

Abstract Generative AI has seen a significant rise in its performance and use in various fields, including academia, art, design, and software engineering. However, little research has been conducted on how users interact with AI tools. This article proposes a research project that focuses on the elicitation of positive psychological experiences, such as flow, when an individual is fully immersed in a task and feels highly effective and satisfied. The study aims to investigate how AI tools may promote flow experiences by regulating cognitive load to flow-conducive levels through providing initial solutions to demanding tasks or shifting user demands from monotonous tasks to more challenging ones. The study will utilize wearable EEG recordings to generate objective insights into the dynamics of cognitive load and flow during early project stages. The research could lead to the development of neuro-adaptive recommender agents that propose AI invocation when undesirable load levels are detected.

Keywords Generative AI · ChatGPT · Cognitive load · Flow · Wearable EEG

1 Introduction

Artificial Intelligence has been on the rise for several years [1, 2]. Yet, with the recent emergence of generalized transformer models like ChatGPT a significant shift in their performance has occurred [3]. In academia, prestigious journals like Nature and Science have already highlighted the rise of Human-AI-collaborative projects [3, 4]. Also, other powerful generative tools for image (e.g. DALL-E), audio (e.g. Speechify), and video generation (e.g. Synthesia), but also coding and software application development (e.g. Uizard) have been developed and impact how people

M. T. Knierim (✉) · L. del Puppo
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: michael.knierim@kit.edu

L. del Puppo
e-mail: lorenzo.puppo@kit.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_6

produce innovative media and knowledge. While some research has focused on the technical capabilities and limitations [5, 6] and socio-economic impacts on the workforce [1, 3], little attention has been paid to the user experience of working with these powerful AI tools. However, as the popularity of ChatGPT has demonstrated (it gained over 100 million users in less than two months), it is crucial to examine how users interact with these tools and the experiences they elicit.

Anticipating a future where people interact with AI tools rather than being replaced by them, it is pertinent to explore how it feels to work with them [7]. This study proposes to focus on one possible user experience outcome: the elicitation of positive psychological experiences [8]. Specifically, we aim to investigate how AI tools may promote experiences of mastery and well-being by facilitating flow, a state of optimal experience characterized by a high level of task engagement and satisfaction [9]. This flow experience has been an emerging topic in IS—and particularly in NeuroIS research (Koufaris, 2002; Léger et al., 2014; Nadj et al., 2023; Knierim et al., 2017; Knierim, 2018; [10–16]). Since flow occurs when skills meet task demands (and when both are present at high levels) [17], and since cognitive load has been identified as a proxy for this skill-demand balance [18], we propose that generative AI tools could regulate cognitive load to flow-conducive levels, for example, by providing initial solutions to challenging tasks (lowering task demands) or by shifting user demands from monotonous tasks to more challenging ones (increasing task demands).

In this article, we propose a study design using NeuroIS methodology to investigate these patterns in a knowledge work task scenario. By using wearable EEG recordings, we also aim to gain objective insights into the dynamics of cognitive load and flow, especially during early project phases. As the EEG is well known for monitoring correlates of cognitive load [19] with high temporal resolution, this approach also allows for theoretical extensions of load-flow dynamics in real-time. In the following sections, we detail the theoretical foundations of flow and cognitive load that we use to develop a corresponding study design. Ultimately, this research could lead to neuro-adaptive recommendation agents that suggest AI invocation when undesirable load levels are detected—and ideally facilitate positive experiences when doing so. Before this happens, however, additional questions need to be answered, which are discussed in the outlook section of this article.

2 Related Work

Flow and Related IS Work. The concept of flow encompasses nine dimensions that include the perception of: (1) a balance between task demand and an individual's skill, (2) clear goals, (3) unambiguous feedback, as well as (4) effortless concentration, (5) merging of action and awareness, (6) loss of self-consciousness, (7) control, (8) distortion of time, and (9) intrinsic reward [9]. Research has shown that flow can be experienced in any task that requires active engagement and fulfills the three preconditions, of which the balance between perceived demand and skill is particularly noteworthy, as it is often manipulated in flow experiments [17]. The concept of flow has been extensively studied in business and IS research, with several studies linking it to improved job performance, creativity, and enjoyment [20–22]. Additionally, it has been associated with technology-enabled team building [23] as well as technology adoption and use [11, 16], highlighting its potential impact on desirable outcomes in the context of IS use and work. The relevance of flow experiences in IS use scenarios and knowledge work has also been recognized in NeuroIS research [10]. While much of the previous research has focused on identifying neurophysiological correlates of flow experiences [14, 15], recent studies have highlighted the connection between flow and cognitive load, which serves as the foundation for this research proposal.

Flow and Cognitive Load. In general, cognitive processing is composed of two primary components: a relatively limited working memory and a much larger long-term memory [24]. When engaging in a task, some degree of working memory must be allocated to that task, a process referred to as cognitive load [24, 25]. Previous work has found that cognitive load and flow experiences are linked through an inverted U-shaped relationship [18, 26]. This has been documented both for self-reports and for well-known EEG correlates of cognitive load like frontal Theta, and posterior Alpha and Beta frequency band power changes [19]. The explanation for this relationship is primarily rooted in the flow pre-condition of demand-skill balance [18]. When a task is neither too demanding nor too easy, the efficient and automated task processing that characterizes flow can occur, likely because processes that are detrimental to the primary task are downregulated (e.g. self-referential attention, conflict monitoring or mind-wandering—[27]). Beyond this general integration of the two theories, little is reported about their cognitive and temporal dynamics. However, this represents an interesting aspect because (1) scholars have reported that flow often emerges sporadically and chaotically [28], and (2) an understanding of the temporal dynamics would provide a highly valuable foundation for the development of flow-facilitating technology (e.g. by understanding when and how to invocate feedback—see [12]). Therefore, in investigating the impact of human-AI collaboration on positive experiences, we propose using NeuroIS methods to study these temporal dynamics to enable theoretical contributions beyond observing their outcomes.

3 Study Proposal

Study Concept, Hypotheses and Procedure. Our proposed study investigates the impact of generative AI tools, such as ChatGPT, on positive experiences in a knowledge work setting. To achieve this, we build on previous NeuroIS research that has examined the emergence of flow experiences in knowledge work tasks like scientific writing [15, 29]. One notable finding from this work was that cognitive load peaked during the start of the writing stage and then decreased after a few minutes [29]. This aligns with flow and writing research that suggests the start of a writing session may require more effort to structure the task than later stages [9, 30]. Based on this observation, we propose that the early stages of a writing project are particularly demanding and that this moment presents an opportunity to use an AI tool to reduce task demands. Thus, we derive the following hypotheses:

H1: Cognitive load increases during early stages of a complex writing task.
H2: Using generative AI reduces cognitive load during early stages of a complex writing task.
H3: The reduction of cognitive load in an early writing stage increases the intensity of flow experiences.

Furthermore, because it is known in flow literature that task skill and importance (here: topic knowledge) significantly influence when and how flow is experienced [31], we expect these two factors to moderate hypotheses H2-H3:

H4a: The more important a topic is to a writer, the greater the AI support's load reduction and flow intensification.
H4b: The more knowledge a writer has with a topic, the weaker the AI support's load reduction and flow intensification.

To account for individual differences in EEG data and skill levels, we propose a fully within-subject experimental design that includes (1) a standardized text copying task of varying difficulty, and (2) drafting two extended abstracts on a standardized topic, once with and once without the support of a generative AI tool. Figure 1 illustrates the proposed experimental structure, including the procedure, tasks, and questionnaires.

Tasks and Manipulations. The text copying task (see, e.g. [32]) is proposed as a baseline measure of cognitive load to resemble common knowledge work. This text copying task consists of a set of pre-selected text segments of varied lengths that are presented to participants for 10 s per segment. By varying the length of each segment

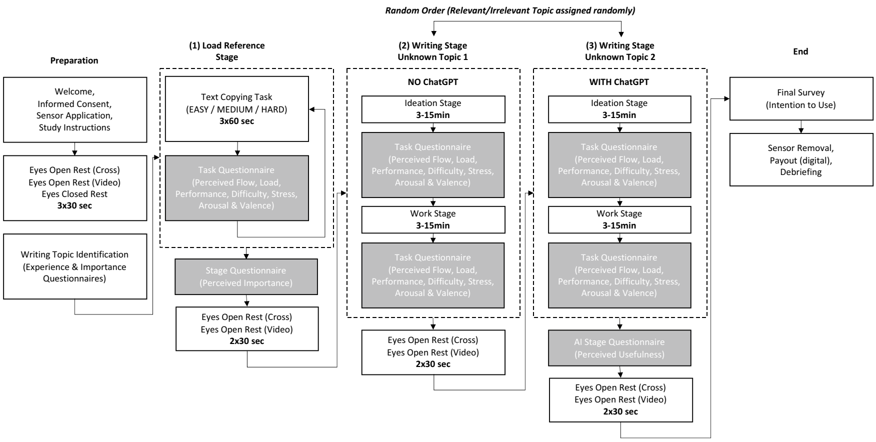


Fig. 1 Procedure of the proposed experiment

that has to be typed, low, medium, and high levels of cognitive load can be elicited that can serve as a reference for the report and EEG data later (see, e.g. [18]).

For the abstract writing stage, participants will work on two standardized topics that are identified in an early stage of the experiment. Specifically, as we expect topic relevance and expertise to moderate the load and flow changes induced by generative AI tools (H4a and H4b), participants are asked to indicate this relevance and expertise for a pre-defined range of four topics at the start of the experiment. These four topics will cover recent developments in the IS discipline. To maximize contrasts, participants will be assigned to work on the topic with the highest and the lowest relevance and expertise—randomly assigned to the AI support/no support conditions. Further, to compare general ability levels for a writing task, we will include the writing practice index (WPI) used in related work on creative writing [33]. The order of the two writing assignments will be randomized to counter any unwanted carryover effects.

The writing process itself is divided into two phases. First, in the ideation stage, participants will have time to draft a storyline for their abstract without the support of any AI tool. They will be asked to focus on generating a story in bullet points. We consider this phase necessary because it will serve to test the hypotheses that cognitive load increases substantially in the early stages of a complex writing task (H1), and that the use of a supportive generative AI tool thereafter reduces cognitive load and flow experience (H2 and H3). In addition, the use of the generative AI tool is more comparable across participants because it is not used for ideation, but rather for implementation of the writing. Second, in the work phase, participants are asked to complete the writing of their text based on their previous ideas. In one round, they are asked to use ChatGPT to produce their extended abstract. In both phases, participants must work for a minimum of three minutes and a maximum of 15 min. We offer this variable time span to account for individual differences in the time it takes participants to produce a satisfactory solution. After each ideation and work stage, participants are asked to fill out surveys about their load and flow experience levels—similar to previous work [15, 34, 35].

To complete the experiment, participants will complete resting baseline measurements for 30 s with eyes open (fixating on a cross) and 30 s watching a video of fish swimming in the ocean as a more natural resting stage stimulus [36]. Overall, we expect the experiment to last 60 min, including the time for instruction and set-up of the instrumentation. Participants will be remunerated with a flat fee following recommendations for flow research [37].

Measures. The NASA Task Load Index (six items by [38]) and the short flow scale (ten items by [31]) will be used as main reports at each task interruption or conclusion. In addition, at each interruption, we will include additional items to assess the task difficulty (one item by [31]) and other facets of the affective experience during the task (a five-item stress measure by [39] and two pictorial items for emotional arousal and valence—[40]). After each task (the three main stages), we will also ask about the task importance (three items by [31]) to compare if task motivation might have influenced the experiences. Finally, at the end of the experiment, participants will

Fig. 2 Dry-electrode wearable EEG system for cognitive load monitoring (see [18])

be asked about their intention to use the AI in similar situations again (three-item TAM construct by [41]—for the AI use condition), or whether they would like to use a generative AI for such tasks (for the non-AI use condition—after description of ChatGPT and its capabilities—adapting the three TAM construct items again).

To continuously monitor cognitive load levels, we will use a wearable EEG system with dry electrodes on the top of the head (importantly with electrodes at Cz)—specifically, a system that resembles headphones that could be used in everyday life (see, e.g. [18, 42] and Fig. 2). We opt for such a system because we believe it is important that NeuroIS scholars engage with such everyday life systems to advance the development of adaptive systems. As cognitive load effects are large and well-observable over frontal and central midline positions [19] we also believe such a system to be an appropriate choice in terms of rigor. We expect to see the classic increase in Theta frequency band power and decrease in Alpha frequency band power as cognitive load increases, especially over the C3, Cz, and C4 positions.

4 Discussion and Outlook

In our work, we propose a study that aims to investigate the impact of generative AI tools on cognitive load and flow experience. We believe that understanding the effects of emerging technologies on user experiences is of paramount importance, especially since the integration of AI tools has the potential to enhance productivity and enjoyment in the future of work through human-AI collaboration. Our proposed study design incorporates wearable EEG systems to observe cognitive dynamics with high temporal resolution, which will contribute significantly to the theoretical integration and extension of load and flow theory. We hope that our findings will provide a novel foundation for explaining how cognitive load patterns contribute to the emergence of flow experiences and how they are regulated by AI tools. It is important to note that our approach is based on the assumption that AI tools provide useful solutions that can be easily integrated into primary tasks. However, if this assumption does not hold, the use of AI tools may have the opposite effect, increasing load and creating more stressful experiences. This possibility highlights the importance of studying the human-AI interaction in knowledge work scenarios, which are anticipated to occur at increasing rates now that generalized transformer models have made their debut in the modern world. As IS scholars, we have an opportunity here to lead the way in studying and supporting the impact of AI on people, organizations, and society as a whole.

References

1. Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254–280.
2. Perotti, J. W., Wall, P., & McLaughlin, G. (2010). The future of knowledge work: Predictions for 2020. *Horiz, 18*, 213–221. https://doi.org/10.1108/10748121011072663
3. Thorp, H. H. (2023). ChatGPT is fun, but not an author
4. Editorial, N. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature, 613*, 612.
5. Dowling, M., & Lucey, B. (2023). ChatGPT for (Finance) research: The Bananarama conjecture. *Finance Research Letters*, 103662. https://doi.org/10.1016/j.frl.2023.103662.
6. Rudolph, J., Tan, S., & Tan, S.: ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*, 23. https://doi.org/10.37074/jalt.2023.6.1.9.
7. Hemmer, P., Westphal, M., Schemmer, M., Vetter, S., Vössing, M., & Satzger, G. (2023). Human-AI collaboration: The effect of AI delegation on human task performance and task satisfaction. In: *IUI '23: Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 453–463). https://doi.org/10.1145/3581641.3584052.
8. Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14. https://doi.org/10.1037//0003-066X.55.1.5
9. Csikszentmihalyi, M. (1996). *Flow and the psychology of discovery and invention*. Harper Collins.
10. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. In: *ACM SIGMIS Database DATABASE Adv. Inf. Syst.* (Vol. 51, pp. 13–54).
11. Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research, 13*, 205–223.
12. Nadj, M., Rissler, R., Adam, M.T.P., Knierim, M.T., Li, M.X., Mädche, A., Riedl, R.: What disrupts flow in office work? A NeuroIS study on the impact of frequency and relevance of IT-mediated interruptions. MIS Q. in press (2023).
13. Léger, P. M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior, 34*, 273–283. https://doi.org/10.1016/j.chb.2014.02.011
14. Knierim, M. T., Rissler, R., Dorner, V., Maedche, A., & Weinhardt, C. (2017). The psychophysiology of flow: A systematic review of peripheral nervous system features. In: *Proceedings of the 9th Retreat on NeuroIS* (pp. 109–120).
15. Knierim, M. T., Rissler, R., Hariharan, A., Nadj, M., & Weinhardt, C. (2018). Exploring flow psychophysiology in knowledge work. In: *Proceedings of the 10th Retreat on NeuroIS* (pp. 239–249).
16. Rissler, R., Nadj, M., & Adam, M. T. P. (2017). Flow in Information Systems research: Review, integrative theoretical framework, and future directions. In: *International Conference on Wirtschaftsinformatik* (pp. 1051–1065).

17. Moneta, G. B. (2021). On the conceptualization and measurement of flow. In C. Peifer & S. Engeser (Eds.), *Advances in flow research* (pp. 31–69). Springer Cham. https://doi.org/10.1007/978-3-030-53468-4_2.
18. Knierim, M. T., Zimny, C., & Seitz, J. (2022). Monitoring cognitive load with 3d-printed EEG headphones. In: *43rd International Conference on Information Systems (ICIS)* (p. 17).
19. Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology,* 1–24.
20. Bakker, A. B., & van Woerkom, M. (2017). Flow at work: A self-determination perspective. *Occupational Health Science, 1*, 47–65.
21. Knierim, M. T., Nadj, M., Li, M. X., & Weinhardt, C. (2019). Flow in knowledge work groups—Autonomy as a driver or digitally mediated communication as a limiting factor? In: *Proceedings of the 40th International Conference on Information Systems (ICIS)* (pp. 1–17).
22. Spurlin, S., & Csikszentmihalyi, M. (2017). Will work ever be fun again? In C. J. Fullagar & A. Delle Fave (Eds.), *Flow at work: Measurement and implications* (pp. 176–187).
23. Keith, M., Anderson, G., Dean, D., Gaskin, J., Wiser, T., Gremmert, S., & Payne, K. (2016). The effects of video gaming on work group performance. In: *Proceedings of the 37th International Conference on Information Systems (ICIS)* (pp. 1–20).
24. Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychology, 38*, 1–4.
25. Grimes, M. G., & Valacich J. S. (2015). Mind over mouse: The effect of cognitive load on mouse movement behavior. In: *Proceedings of the 36th International Conference on Information Systems (ICIS)* (pp. 1–13).
26. Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop. *Frontiers in Human Neuroscience, 10*, 1–13.
27. Harris, D. J., Vine, S. J., & Wilson, M. R. (2017). Neurocognitive mechanisms of the flow state. *Progress in Brain Research, 237*, 221–243. https://doi.org/10.1016/bs.pbr.2017.06.012
28. Ceja, L., & Navarro, J. (2012). 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Hum. Relations., 65*, 1101–1127. https://doi.org/10.1177/0018726712447116
29. Knierim, M. T., Nadj, M., Hariharan, A., & Weinhardt, C. (2018). Flow neurophysiology in knowledge work: Electroencephalographic observations from two cognitive tasks. In: *5th International Conference on Physiological Computing Systems (PhyCS)* (pp. 42–53).
30. Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*, 365–387.
31. Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*, 158–172.
32. Fest, S., Kvaløy, O., Nieken, P., & Schöttner, A. (2021). How (not) to motivate online workers: Two controlled field experiments on leadership in the gig economy. *The Leadership Quarterly, 32*, 16.
33. Lotze, M., Erhard, K., Neumann, N., Eickhoff, S. B., & Langner, R. (2014). Neural correlates of verbal creativity: Differences in resting-state functional connectivity associated with expertise in creative writing. *Frontiers in Human Neuroscience, 8*, 516. https://doi.org/10.3389/fnhum.2014.00516
34. Knierim, M. T., Nadj, M., Hariharan, A., & Weinhardt, C. (2018). Flow neurophysiology in knowledge work: Electroencephalographic observations from two cognitive tasks. In *Proceedings of the 5th International Conference on Physiological Computing Systems (PhyCS)* (pp. 42–53).
35. Rissler, R., Nadj, M., Li, M. X., Loewe, N., Knierim, M. T., & Maedche, A. (2020). To be or not to be in flow at work: Physiological classification of flow using machine learning. *IEEE Transactions on Affective Computing*, 1–12.
36. Piferi, R. L., Kline, K. A., Younger, J., & Lawler, K. A. (2000). An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video. *International Journal of Psychophysiology, 37*, 207–217.

37. Moller, A. C., Meier, B. P., & Wall, R. D. (2010). Developing an experimental induction of flow: Effortless action in the lab. In B. Bruya (Ed.), *Effortless attention* (pp. 191–204). https://doi.org/10.7551/mitpress/9780262013840.003.0010.
38. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.
39. Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., & Grover, V. (2014). NeuroIS—Alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems, 15*, 723–753.
40. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59.
41. Davis, F. D. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results.
42. Kartali, A., Janković, M. M., Gligorijević, I., Mijović, P., Mijović, B., & Leva, M. C. (2019). Real-time mental workload estimation using EEG. In: *International Symposium on Human Mental Workload: Models and Applications* (pp. 20–34).

Investigating the Impact of Emotions on the Quality of Novice Programmers' Code

Lisero Mugula, Rosemary Tufon, Yang Shi, Maria Valero, Adriane B. Randolph, and Valentina Nino

Abstract This paper examines the impact of emotions on the quality of code for novice programmers while developing in an experimental setting using Python. A relaxation mechanism was additionally-used to explore the possible effect on emotion and programmer performance. Non-invasive EEG was recorded to assess activations in the left and right prefrontal cortex often related to emotions. The quality of code was obtained based on the Code-based Deep Knowledge Tracing method. Contrary to expectations, preliminary results show that positive emotions may contribute to novice programmers generating lower-quality code. Extension of this study may help solidify this relationship while reinforcing the corollary of how negative emotions contribute to code quality and recommendations for how to situate programmers in these varying mental states.

Keywords Emotions · EEG · Quality of programming code · Novice programmers · Relaxation

L. Mugula · R. Tufon · M. Valero · A. B. Randolph (✉) · V. Nino
Kennesaw State University, Kennesaw, Georgia, USA
e-mail: arandol3@kennesaw.edu

L. Mugula
e-mail: lmugula@students.kennesaw.edu

R. Tufon
e-mail: rtufon@students.kennesaw.edu

M. Valero
e-mail: mvalero2@kennesaw.edu

V. Nino
e-mail: lvallad1@kennesaw.edu

Y. Shi
North Carolina State University, Raleigh, North Carolina, USA
e-mail: yshi26@ncsu.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_7

1 Introduction

Emotion is one of the most popular areas of focus for the neuroIS community where neurophysiological tools may be used for more objective measures than self-reports [1, 2]. Emotion is characterized in psychology as a multidimensional state where an individual experiences feelings that can be described as good or bad, energizing or draining. A growing body of literature is recognizing a two-dimensional framework, the circumplex model, for representing this phenomenon. This model describes an emotional state in a two-dimensional circular space with the arousal and valence dimensions, where the dimension of valence describes the extent to which an emotion is positive or negative, and the dimension of arousal describes its intensity [3–5]. These emotional experiences have been shown to be associated with activation of certain regions in the brain suggesting that physiological correlates of emotions can be measured using neurophysiological and neuroimaging tools [6, 7].

Electroencephalography (EEG), which measures electrical activity in the brain, is one of such neurophysiological measures that correlate with emotions [2, 8]. Changes in emotions can cause differences in EEG signals, which may reflect different emotional states. For instance, research shows that high frequencies sensed from electrodes positioned on the frontal, parietal, and occipital lobes have high correlation with emotional valence [9], and it is possible to discriminate between happiness and sadness by analyzing EEG signals in various situations [10]. In fact, EEG has been used to study how the brain functions in relation to job performance in software development and how programmers experience positive and negative emotions when coding.

Code quality is an important metric for evaluation of software as poorly written code can negatively impact the utility and the security of an application. The quality of code refers to how well-written and structured the code is, as well as how effectively it performs its intended function [11, 12]. High-quality code is easy to read and understand, well-documented, and efficient. There are many technical factors that affect the quality of code, such as clarity, maintainability, reliability, and security [13]. Other behavioral factors can impact the quality of code, such as motivation, communication, time management, and emotions [14].

Emotions can have a significant impact on people's performance. Notably, positive and negative emotions can significantly impact the quality of code that developers produce [15–18]. Relaxation can have a positive impact on code quality, as well. When developers are relaxed, they are more likely to be focused, creative, and attentive to detail [19–21]. These characteristics can lead to better code quality and more efficient development.

In this paper, we present a work-in-progress study to understand the impact of emotions on the quality of programming code. To better understand this impact, we explore the possible effect of relaxation on emotions while novice programmers solve a case problem. We provide a discussion about these findings with appreciation that further investigation with a larger sample size will be most helpful for the continuation of this research to reinforce or alter current leanings.

2 Background

Despite growing evidence that most defects in code quality are a result of human cognitive issues [22–25], software development processes in practice do not consider the emotional state of programmers during coding. The ability to examine and understand how emotions affect code quality using neurophysiological measurements can provide guidance to managers and developers to take appropriate measures to minimize and prevent poor-quality code and promote a healthy environment for the programmer. Furthermore, scholars have suggested that the use of neurophysiological measures in software engineering could lead to the design and development of integrated, neuro-adaptive development environments and learning platforms [26–29]. Several studies in software engineering have employed a variety of neurophysiological tools to study the electrical activity of the brain [2, 8], the skin [28], and heart rate variability (HRV) [29]. Results from these studies indicate that measures captured with neurophysiological tools can be used to investigate the emotional states of programmers when performing various programming tasks [3]. However, most of this prior work has centered on the prediction of abilities. Only a limited number of studies focus on emotional processes using neurophysiological tools in the context of software engineering [30–32]. Our study builds on prior work by using neurophysiological tools to examine the impact of relaxation and emotions on the quality of code produced by novice programmers during a coding task.

3 Proposed Study

Electroencephalography has been used to extract and detect positive and negative emotions from brain signals and incorporated as feedback into neuro-adaptive interfaces [33, 34]. The proposed study aims to use non-invasively recorded EEG to measure developers' emotions while programming a task to identify if positive and negative emotions have a correlation with high- or low-quality of the code. The EEG recording system we are using is the BioSemi ActiveTwo bioamplifier system [35].

The ActiveTwo system consists of a headbox, which connects to a set of electrodes placed on the scalp to measure EEG signals, and a bioamplifier unit, which magnifies and digitizes these signals. The ActiveTwo system is used to measure EEG activity in specific regions of the brain that are associated with emotional processing, such as the prefrontal cortex [36]. The prefrontal cortex is a brain region involved in many cognitive and emotional processes, including the regulation of positive and negative emotions [37]. It has been shown that the left prefrontal cortex (PFC) is associated with positive emotions, while the right PFC is associated with negative emotions [38–40].

The quality of code can be determined using Code-DKT [41]. Code-DKT stands for "Code-based Deep Knowledge Tracing." It is a neural network-based approach to modeling student knowledge over time in the context of programming tasks.

In this study, we use the categories of Code-DKT to manually analyze the code's quality, specifically on the reason why participants make incorrect submissions. In the future, we plan to use Code-DKT in a more data-driven way to evaluate novices' code quality. For example, Code-DKT uses the Code2vec model to predict students' programming performance in an educational context based on their submissions when they practice programming in a course [42]. The approach incorporates the Code2vec model into the traditional DKT model for the performance prediction task [43]. This approach directly incorporates features from programming code into the model and achieves state-of-the-art performance in the prediction task for the computing education domain. While we currently leverage a limited set of test cases to evaluate the quality of programming code, as more data is available, we will switch to the data-driven method for a more accurate evaluation of the quality of student code quality.

4 Methodology

Seven (7) participants were recruited with ages between 18 and 40 years old and located in the southeast region of the United States. Participants belonged to different university departments with some fundamental coding experience with Python. Participants were randomly divided into two groups. The control group received a relaxation treatment via a biofeedback device called the Muse 2 [44] for ten minutes. The experimental group attempted the coding task without engaging in the relaxation activity.

The BioSemi ActiveTwo system was used to collect participants' EEG signals as they completed a thirty-minute coding activity of writing Python code for a classical programming problem, the rainfall problem [45]. The placement of the BioSemi ActiveTwo sensors was performed by trained personnel to ensure that it was comfortably and accurately positioned on the participant's head. The study ensured that the procedures for both, the experimental and control groups, were consistent and standardized to reduce any potential confounding factors that may affect the study results. Figure 1 shows one of the participants wearing an electrode cap fitted with active electrodes wired to the BioSemi bioamplifier during the experiment.

Independent variables consisted of individual characteristics, including age, sex, relaxation treatment, and EEG data from the brain that was collected during the time the participants attempted the coding task. We collected information from 16 channels of EEG using a Common Average Reference (CAR) across: frontal-polar (Fp1, Fp2), frontal-central (FC3, FCz, FC4), central (C3, Cz, C4), temporal-parietal (TP7, TP8), parietal (P3, Pz, P4), and occipital (O1, Oz, O2). After collection, we performed offline feature extractions of the EEG signals for emotion classification. We used the software EDFBrowser [46] to down sample the signals from 1024 Hz to a sampling rate of 256 Hz. Then, we used standardized low resolution brain electromagnetic tomography (sLORETA) software to determine the cross-spectra and visualize the

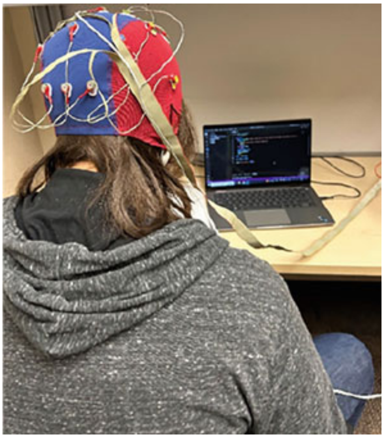


Fig. 1 Participant wearing an electrode cap with BioSemi active electrodes

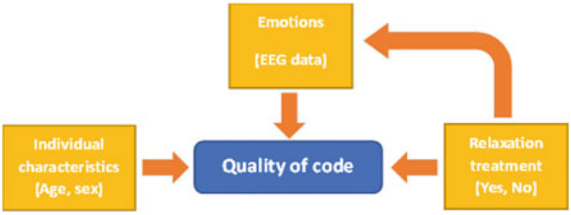


Fig. 2 Independent variables of age, sex, relaxation treatment, and emotion as determined from EEG, and their possible effects on the independent variable of code quality

brain activations in a topographic plot [47]. The topographies revealed the left and right prefrontal cortex asymmetry as often related to emotions.

Dependent variables consisted of the manual analysis of the code's quality that consisted of parameters such as syntaxis errors, I/O incorrect prompt, and non-identifiable logic. Quality points were assigned from 0 to 4, where 0 is the lowest quality and 4 is the highest. Figure 2 shows a representation of the possible relationships between the variables of the study. Independent variables such as age, sex, and relaxation treatment may act as moderators or mediators on the quality of code, however a bigger sample size is needed to establish such effects.

5 Preliminary Results

Emotions and code quality were analyzed for the seven participants in the study of six males and one female with an average age of 21.6 (±3.7). The participants were racially-diverse, with two White, three Asian, one Hispanic, and one Black/African-American. After visual inspection of the EEG, we analyzed the brain activations from the sixteen channels of scalp electrodes using a previously-validated technique for brain localization with the same name of the accompanying software, sLORETA [48]. These activations were presented on a normalized scale such that brighter areas with yellow indicated the highest levels of activation. For each six-view grouping of topological plots, the image presented on the top row in the center is a back-end view of the brain whereas the image on the bottom row in the center is a front-on view of the brain. In a similar manner as used by Riley and Randolph [49], theory indicates that higher activation in the left hemisphere may indicate a stronger positive approach to the stimulus whereas higher activation in the right hemisphere may indicate a negative approach to the stimulus [49, 50].

As "work-in-progress," we start seeing a tendency in the results that need to be confirmed with more data. Table 1 shows the results of each participant after completing the study. The Muse 2 provided an assessment of percent calm which we report for all four participants who received the relaxation treatment except for participant C0110 whose percentage was not captured.

Surprisingly, the majority of the participants that experienced positive emotions (evidenced from highest activation levels in left prefrontal cortex), developed low quality code. There are inconclusive results from participants that exhibited negative emotions (evidenced from highest activation levels in right prefrontal cortex) and their quality of code where there is an even split between high and low quality of code generated.

Figure 3 shows exemplar topography of brain activation for participant C0300. As can be seen in Fig. 3, the yellow color represents highest levels of brain activation that were observed in the right prefrontal cortex and has been related to negative emotions. Additionally, the high activity was in the Brodmann Area 45 [51], which has been associated with semantic tasks, such as semantic decision tasks (determining whether a word represents an abstract or a concrete entity) and generation tasks. This participant ranked three out of four in the analysis of code quality, which means that overall, the code produced the expected result with minimal error of a division by zero. In contrast as seen in Fig. 4, participant A200 exhibited positive emotions while the quality of the code was low. Relaxation seemed to have a contrary effect where participants who received the relaxation treatment seemed to exhibit negative motions.

Table1 Preliminary results of the study

ID	Emotion from EEG	Quality of code	Relaxation (% Calm)
A100	Positive	0/4	–
A200	Positive	0/4	–
A500	Positive	3/4	–
C0110	Negative	4/4	Y (NA)
C0200	Negative	0/4	Y (70%)
C0300	Negative	3/4	Y (42%)
C0900	Negative	0/4	Y (24%)

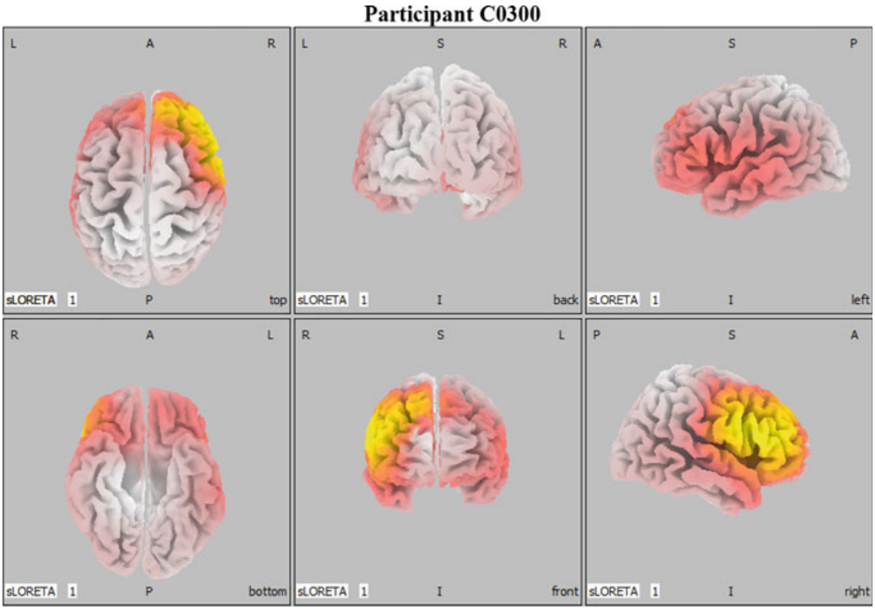


Fig. 3 Topography of brain activation for Participant C0300

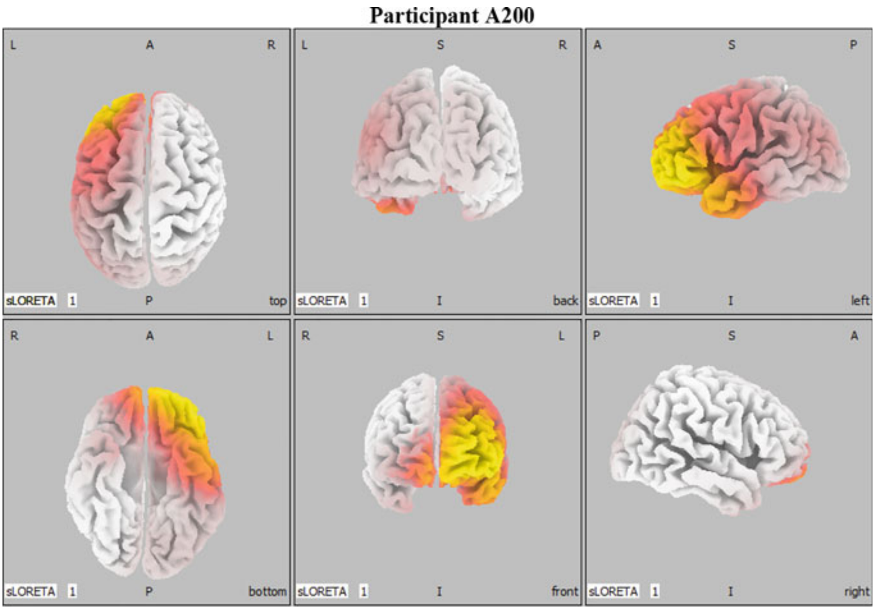


Fig. 4 Topography of brain activation for Participant A200

6 Discussion

Emotions are not the only factor that influences the quality of code. Other important factors also have impact, such as developer's skill, confidence, and motivation. In this preliminary study, we showed that positive emotions seem to favor the lower quality of code. This statement seems to be alienated with general literature indicating that stronger positive emotions are correlated with higher confidence in successful task completion and complicated with how high-performance goal orientation is related to stronger negative and weaker positive emotions [52]. While surprising, these preliminary results encourage more discussion and further investigation.

A deep discussion about the preliminary results led to potential research questions that can be explored in future work. For example, looking at the initial results, we can see that everyone who obtained the relaxation treatment had negative emotions, but some of them had good-quality code. This requires more investigation, keeping in mind the ones developing the code are humans that get affected by their emotional load. Some practitioners have found that happiness, a positive emotion, could lead to more code (quantity) while sadness, a negative emotion, leads to better and cleaner code (fewer mistakes) [53]. Some others have found that stress, anger, arrogance, and negative emotions lead to more code errors and in contrast motivation and confidence, positive emotions, lead to better code or code that is easy to read and maintain [54]. Therefore, the results we are getting so far might not be entirely contradictory with literature.

Our results do, however, indicate the need to perform a more in-depth analysis about the specific positive and negative emotions experienced by the participants to understand the relationship between emotions and quality of code. Likewise, due to the highly diverse nature of the participants, it is also possible that the effects were due to differences in expertise rather than the affective state. For this reason, we propose in future work to have a self-reporting evaluation of the level of expertise of the participants. Other potential confounds need to also be considered.

7 Future Work

As next steps, we propose a research roadmap to explore more emotional factors on the quality of code. First, we will design a separate study with around forty participants and one more complex programming code task with a prior self-report of expertise in the programming language to help avoid confounding factors affecting the results. We will analyze participants in different relaxation scenarios to measure the impact of relaxation. We will compute results to verify the correlation between emotions and code quality and reduce inconclusive results. Finally, we will introduce and evaluate the impact of HRV on emotions and code quality, as other literature has shown that this is a significant measure [55, 56]. In this roadmap, it is also essential to highlight that for the preliminary results, the relaxation device (Muse2) is not considered a research-grade tool. Due to the weighted discussion of the use of research-grade devices [57, 58], we may find a substitute for measuring the relaxed state for individuals.

8 Conclusion

In this paper, we show an experimental work-in-progress for determining the impact of emotions on the quality of code of novice programmers. We designed a study that non-invasively obtained EEG signals from seven participants while they worked on developing a programming task. According to the preliminary results, positive emotions may lead to lower-quality code. Further investigation with a larger sample size is needed to validate this direction.

References

1. Riedl, R., Fischer, T., Léger, P. M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 51*(3), 13–54.
2. Walla, P. (2017). Affective Processing Guides Behavior and Emotions Communicate Feelings: Towards a Guideline for the NeuroIS Community. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Lecture Notes in Information Systems and Organisation: Vol. 25. Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017* (pp. 141–150). Springer Berlin Heidelberg.
3. Girardi, D., Novielli, N., Fucci, D., & Lanubile, F. (2020, June). Recognizing developers' emotions while programming. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering* (pp. 666–677).
4. Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. *Attention and Orienting: Sensory and Motivational Processes, 97*, 135.
5. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145.
6. Russell, J. A., Ward, L. M., & Pratt, G. (1981). Affective quality attributed to environments: A factor analytic study. *Environment and Behavior, 13*(3), 259–288.
7. Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209–237.
8. Rubin, D. C., & Talarico, J. M. (2009). A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory, 17*(8), 802–808.
9. Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2015). Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing, 7*(1), 17–28.
10. Li, M., & Lu, B. L. (2009, September). Emotion classification based on gamma-band EEG. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1223–1226). IEEE.
11. Kim, M. K., Kim, M., Oh, E., & Kim, S. P. (2013). A review on the computational methods for emotional state estimation from the human EEG. *Computational and Mathematical Methods in Medicine.*

12. Fitzpatrick, R. (1996). *Software quality: Definitions and strategic issues*. Staffordshire University.
13. Kononenko, O., Baysal, O., & Godfrey, M. W. (2016, May). Code review quality: How developers see it. In Proceedings of the 38th international conference on software engineering (pp. 1028–1038).
14. Günsel, A. (2014). The effects of emotional labor on software quality: The moderating role of project complexity. *Journal Of Global Strategic Management, 8*(2), 101–115.
15. Dahlberg, D. (2020). Developer experience of a low-code platform: An exploratory study.
16. McCarthy, P. J. (2011). Positive emotion in sport performance: Current status and future directions. *International Review of Sport and Exercise Psychology, 4*(1), 50–69.
17. Graziotin, D., Fagerholm, F., Wang, X., & Abrahamsson, P. (2018). What happens when software developers are (un) happy. *Journal of Systems and Software, 140*, 32–47.
18. Girardi, D., Lanubile, F., Novielli, N., & Serebrenik, A. (2021). Emotions and perceived productivity of software developers at the workplace. *IEEE Transactions on Software Engineering, 48*(9), 3326–3341.
19. Moster, M., Chandra, A., Chu, C., Liu, W., & Rodeghero, P. (2022, September). In the zone: An analysis of the music practices of remote software developers. In *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 313–318).
20. Affaldano, G., Steinmacher, I., Calefato, F., Gerosa, M., & Lanubile, F. (2019). Why do developers take breaks from contributing to OSS projects? A preliminary analysis. arXiv preprintarXiv:1903.09528.
21. Volobuev, V., Turischeva, P., Gainutdinov, A., Sahibgareev, R., Mazzara, M., & Farina, M. (2021, August). Effects of Mindfulness meditation on software developers' performance. In *2021 International Conference "Nonlinearity, Information and Robotics" (NIR)* (pp. 1–6). IEEE.
22. Basile, V., Novielli, N., Croce, D., Barbieri, F., Nissim, M., & Patti, V. (2018). Sentiment polarity classification at evalita: Lessons learned and open challenges. *IEEE Transactions on Affective Computing, 12*(2), 466–478.
23. Bradley, M. M., Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavioral Therapy and Experimental Psychiatry, 25*, 49–59.
24. Graziotin, D., Wang, X., & Abrahamsson, P. (2014). Software developers, moods, emotions, and performance. arXiv preprintarXiv:1405.4422.
25. Müller, S. C., & Fritz, T. (2015, May). Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering* (Vol. 1, pp. 688–699). IEEE.
26. Weber, B., Fischer, T., & Riedl, R. (2021). Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review. *Journal of Systems and Software, 178*, 110946.
27. Adam, M. T., Gimpel, H., Maedche, A., & Riedl, R. (2017). Design blueprint for stress-sensitive adaptive enterprise systems. *Business & Information Systems Engineering, 59*, 277–291.
28. Burleson, W., & Picard, R. W. (2004, August). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In *Workshop on Social and Emotional Intelligence in Learning Environments*.
29. Canento, F., Fred, A., Silva, H., Gamboa, H., & Lourenço, A. (2011, October). Multimodal biosignal sensor data handling for emotion recognition. In *SENSORS, 2011 IEEE* (pp. 647–650). IEEE.
30. Müller, S. C., & Fritz, T. (2016, May). Using (bio) metrics to predict code quality online. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 452–463).
31. Vrzakova, H., Amon, M. J., Stewart, A., Duran, N. D., & D'Mello, S. K. (2020, March). Focused or stuck together: multimodal patterns reveal triads' performance in collaborative problem solving. In *Proceedings of the Tenth International Conference On Learning Analytics & Knowledge* (pp. 295–304).

32. Sánchez-Gordón, M., & Colomo-Palacios, R. (2019). Taking the emotional pulse of software engineering—A systematic literature review of empirical studies. *Information and Software Technology, 115*, 23–43.
33. Kaur, B., Singh, D., & Roy, P. P. (2018). EEG based emotion classification mechanism in BCI. *Procedia Computer Science, 132*, 752–758.
34. Constant, I., & Sabourdin, N. (2012). The EEG signal: A window on the cortical brain activity. *Pediatric Anesthesia, 22*(6), 539–552.
35. BioSemi ActiveTwo system website: https://www.biosemi.com/products.htm.
36. Dalgleish, T. (2004). The emotional brain. *Nature Reviews Neuroscience, 5*(7), 583–589.
37. Carlén, M. (2017). What constitutes the prefrontal cortex? *Science, 358*(6362), 478–482.
38. Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences, 15*(2), 85–93.
39. Davidson, R. J. (2004). What does the prefrontal cortex "do" in affect: Perspectives on frontal EEG asymmetry research. *Biological Psychology, 67*(1–2), 219–234.
40. Gotlib, I. H. (1998). EEG alpha asymmetry, depression, and cognitive functioning. *Cognition & Emotion, 12*(3), 449–478.
41. Shi, Y., Chi, M., Barnes, T., & Price, T. (2022). Code-DKT: A Code-based Knowledge Tracing Model for Programming Tasks. In *Proceedings of the 15th International Conference on Educational Data Mining* (p. 50).
42. Alon, U., Zilberstein, M., Levy, O., & Yahav, E. (2019). code2vec: Learning distributed representations of code. In *Proceedings of the ACM on Programming Languages, 3(POPL)* (pp. 1–29).
43. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems, 28.*
44. Muse 2. The brain sensing headband. https://choosemuse.com/muse-2/.
45. Seppälä, Otto, et al. (2015). Do we know how difficult the rainfall problem is? *Proceedings of the 15th Koli Calling Conference on Computing Education Research.* https://dl.acm.org/, https://doi.org/10.1145/2828959.2828963.
46. EDFBrowser. https://www.teuniz.net/edfbrowser/.
47. SLoreta: Low Resolution Brain Electromagnetic Tomography. https://www.uzh.ch/keyinst/loretaOldy.htm.
48. Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology, 24*(Suppl D), 5–12.
49. Riley, J., & Randolph, A. B. (2021). Leveraging NeuroIS Tools to Understand Consumer Interactions with Social Media Content. In *Information Systems and Neuroscience: NeuroIS Retreat 2021* (pp. 56–62). Springer International Publishing.
50. Davidson, R. J. (1992). Anterior Cerebral Asymmetry and the Nature of Emotion. *Brain and Cognition, 20*, 125–151.
51. Strotzer, M. (2009). One century of brain mapping using Brodmann areas. *Clinical Neuroradiology, 19*(3), 179.
52. Fisher, C. D., Minbashian, A., Beckmann, N., & Wood, R. E. (2013). Task appraisals, emotions, and performance goal orientation. *Journal of Applied Psychology, 98*(2), 364.
53. Ziegler, A. (November 2018). Decoding programmers: How emotions can change code. Information Week. https://www.informationweek.com/devops/decoding-programmers-how-emotions-can-change-code.
54. Gregory, K. (Nov-Dec 2019). Emotional code. Code Magazine. https://www.codemag.com/article/1911081/emotional-code.
55. Hermes, A., & Riedl, R. (2021). Exploring the influence of personality traits on affective customer experiences in retailing: combination of heart rate variability (HRV) and self-report measures. In *Information Systems and Neuroscience: NeuroIS Retreat 2021* (pp. 20–29). Springer International Publishing.
56. Park, S. S., & Lee, K. C. (2019). Implementing an adaptive neuro-fuzzy model for emotion prediction based on heart rate variability (HRV). *Journal of Digital Convergence, 17*(1), 239–247.

57. Müller-Putz, G. R., Tunkowitsch, U., Minas, R. K., Dennis, A. R., & Riedl, R. (2021). On electrode layout in EEG studies: A limitation of consumer-grade EEG instruments. In *Information Systems and Neuroscience: NeuroIS Retreat 2021* (pp. 90–95). Springer International Publishing.
58. Riedl, R., Minas, R. K., Dennis, A. R., & Müller-Putz, G. R. (2020). Consumer-grade EEG instruments: Insights on the measurement quality based on a literature review and implications for NeuroIS research. *Information Systems and Neuroscience: NeuroIS Retreat, 2020*, 350–361.

Does Self-View Mode Generate Video Conferencing Fatigue? An Experiment Using EEG Signals

Jin Xu, Eoin Whelan, Ann O'Brien, and Denis O'Hora

Abstract The ability to see or hide one's own image is a typical feature of video conferencing platforms. This study will conduct an EEG-based neurobiological experiment to determine if the self-view mode generates video conference fatigue and if this differs between males and females. 40 volunteers will participate in a simulated video conference meeting with the self-view mode on and off at different times. In addition, an EEG-based fatigue monitor will be proposed to demonstrate the level of human mental fatigue. The experimental insights will provide direct biological evidence of the impact of video conferencing features on the user experience and these will be of benefit to inform the design of web conferencing platforms and improve the user experience of video conferencing.

Keywords Video conference · Fatigue measurement · Self-view · EEG analysis

1 Introduction

The Covid-19 pandemic has forced a dramatic increase in the number of video conferencing sessions for work purposes. In a post-pandemic world, video conferencing solutions will remain central as organisations continue to support hybrid work options. Some studies [1–10] have shown that engaging with certain video

J. Xu (✉) · E. Whelan · A. O'Brien
J.E. Cairnes School of Business & Economics/Insight Centre for Data Analytics, University of Galway, Galway, Ireland
e-mail: xuj@universityofgalway.ie

E. Whelan
e-mail: eoin.whelan@universityofgalway.ie

A. O'Brien
e-mail: ann.obrien@universityofgalway.ie

D. O'Hora
School of Psychology, University of Galway, Galway, Ireland
e-mail: denis.ohora@universityofgalway.ie

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_8

conferencing platform features can trigger fatigue, such as self-view mode. Some theoretical analysis indicated that being stared at was a significant predictor of Video Conference Fatigue (VCF) when users look at their screens leading to increased self-awareness. However, there are still no biological experiments to directly demonstrate the effect of self-view mode on VCF. Furthermore, many previous studies have simulated video conferencing scenarios through volunteers watching a series of videos on the computer, but this does not realistically simulate video conferencing in our opinion. In this study, we will use the self-view mode in Zoom to investigate the effect of the self-view on VCF in a real video interview scenario. A novel EEG-based VCF analysis framework will be presented and an EEG fatigue monitor will be demonstrated to show the level of mental fatigue which involves 40 volunteers. The experiment results can give us biological evidence to identify if turning on/off self-view mode can reduce the VCF. Furthermore, the effect of the self-view mode on VCF for males and females of different genders will be analysed. These findings can also inform the design of video conferencing platforms to limit the negative impacts on user well-being.

2 Problem Statement

EEG signals for VCF analysis are used to build on our insights by precisely determining how engagement with the video conferencing feature self-view mode affects user fatigue levels and if this differs between males and females. The option of self-view mode is supported on Zoom. One of the options in self-view is called "Show Self View" in Zoom, it allows the user to view themself. Another one is self-view mode off called "Hide Self View" in Zoom, it allows users to hide the video of themself from their own screen, even though others in the meeting can see their video. Recent research revealed that the self-view mode can affect the VCF through increased self-awareness and disrupts the automatic processes that are typical for effective communications [1, 6, 8, 9, 21]. While some research has investigated VCF using questionnaires in distance learning, there are still no biological experiments investigating the effect of self-view mode on mental fatigue in a real video conferencing scenario. Therefore, in this study, an EEG experiment will be conducted where the volunteer's EEG will be acquired in a real online interview conversation scenario using Zoom under the self-view mode on and off.

3 Related Work

Video Conference Fatigue

Due to the massive global use of video conferencing tools for simultaneous remote communication over the past two years, more and more people are experiencing symptoms of mental and physical fatigue. VCF is defined as somatic and cognitive exhaustion that is caused by the intensive and/or inappropriate use of videoconferencing tools, frequently accompanied by related symptoms such as tiredness, worry, anxiety, burnout, discomfort, and stress, as well as other bodily symptoms such as headaches [1]. In 2020, Morris demonstrated how mental fatigue is related to VCF and what are the causes and dynamics [2]. Mainly this is caused by exhaustion with online communication. Following the pandemic enforced lockdown and social distancing, where people have been connected using an online mode of communication, this type of mental fatigue has increased. Nadler has discussed the causes of VCF, from the online mode of communication, and the effect of cognitive load on individuals [3]. Fauville et al. used a series of surveys to measure video conferencing fatigue and indicated that frequency, duration, and burstiness of Zoom meetings were associated with a higher level of fatigue [4]. In 2021, Massner presented multi-dimensional factors that lead to VCF, such as the number of video conferences scheduled a day, the size of the video conference, the relationship among participants, the type of content shared in the video conference, the level of participation (host or participant), and the amount of interaction during the video conference [5]. In 2022, Li et al. summarised that factors causing VCF include unnatural interaction with multiple faces mental fatigue detection, self-view, asynchronicity, lack of body language, lack of eye contact, cognitive load, multitasking and reduced mobility [6]. In a study involving 33 volunteers the associations between video conference fatigue, burnout, depression and personality trait neuroticism were investigated and the study indicated that these four constructs were robustly positively associated with each other [7]. Theoretical analysis indicates that if a user's own face is shown on the interface, it may result in more pronounced perceptions of cognitive exhaustion and fatigue, due to increased attentional and working memory demands [1, 8]. Differences in fatigue by gender of video conferencing participants when they look at their screen have been identified, leading to females experiencing greater Zoom fatigue than males [9]. In this work, a neurophysiological experiment will be designed by using EEG signals to detect human mental fatigue on a video conference with the on/off self-view at different times. EEG signals can directly respond to human fatigue levels and will provide biological evidence to demonstrate how the self-view model affects VCF and to verify the impact of gender on VCF.

EEG Fatigue Measurement

In general, EEG signals are closely related to mental fatigue [10]. When large numbers of nerve cell groups are synchronised, EEG signals can record changes in postsynaptic potentials for analysis and research [11]. EEG signals have been used to detect mental fatigue in humans. In the work of Acı et al. [12], some machine learning algorithms were used for mental fatigue detection. In the work of Deng et al. [11], EEG signal provides four basic fatigue indicators. During fatigue, the slow wave increases while the fast wave decreases accordingly. At the same time, the powers of δ and θ increase, while the powers of α and β decrease. In the work of Abdulhamit [13], they indicated that within NREM sleep, δ power (slow wave power) indicates the intensity of sleep. In the work of Saroj et al. [14], they proposed an algorithm for detecting different levels of fatigue and FFT was used to transform raw EEG data into the frequency domain. In the work of Jap et al. [15], they used four algorithms for fatigue detection, which were: $(\theta + \alpha) / \beta$, α / b, $(\theta + \alpha) / (\alpha + \beta)$ and θ / β, were also assessed as possible indicators for fatigue detection. In the work of Simon et al. [16], a method for extracting EEG α spindles under noisy recording conditions was presented. In the real road driving experiment, α spindle measures could reliably identify driver fatigue and clearly differentiate between fatigue and time-on-task effects. In the work of Trejo et al. [17], they indicated that Mental fatigue was associated with increased power in frontal θ and parietal α EEG rhythms. A statistical classifier can use these effects to model EEG-fatigue relationships accurately. In the work of Ashley Craig et al. [18], they showed that as an individual grows fatigued, slow wave activity such as θ and α activity increases over the entire cortex. The results showed that as a person fatigues, slow wave activity increased over the entire cortex, in θ and in $\alpha 1$ and 2 bands, while no significant changes were found in δ wave activity. Table. 1 summarises the research on EEG-based fatigue analysis. It can be found that the main method of analysis is using EEG spectral information, for example using the power ratio between different EEG frequency bands and other variants (e.g. α Spindle Rate). Another option is to use classification methods to train machine learning models to detect mental fatigue.

4 Methodology

Volunteers and Task

To achieve our research goal, 40 volunteers will be recruited for this study. Before commencing the experiments, volunteers will complete a short survey e.g., age, gender, and video conferencing experience. Volunteers conduct two real video interview sessions under self-view mode on and self-view model off and their EEG data will be collected simultaneously. To consider possible order effects, these volunteers will be divided into two groups. The first group will have the first half of their video

Table 1 Summary of studies for EEG-based fatigue measurement

Research work	Channels	Sampling frequency	Spectral extraction	Mental state classification
Deng et al. [11]	64	160 Hz	FFT	$(\delta + \theta) / (\alpha + \beta)$ + DCSAEN
Acı et al. [12]	7	128 Hz	STFT	SVM
Abdulhamit [13]	8	150 Hz	DWT	ANN
Saroj et al. [14]	19	256 Hz	FFT	Lab view tool
Jap et al. [15]	30	1000 Hz	FFT	$(\theta + \alpha) / \beta$, α / β, $(\theta + \alpha) / (\alpha + \beta)$, and θ / β
Simon et al. [16]	128; 64	1000 Hz; 500 Hz	STFT	α spindle rate
Trejo et al. [17]	32	500 Hz	DWT-8	Linear regression classifier
Ashley et al. [18]	32	1025 Hz	FFT	Chalder Fatigue Scale (CFS)

interview in self-view mode on (20 min), have 10 min break, and then participate in the second half via self-view mode off (additional 20 min). The other group of students will have the first half of the video interview via self-view mode off, have a 10 min break, and then participate in the second half via self-view on. Based on this procedure, possible carry-over effects can be considered in statistical analyses. The gender of each group subject will be half women and half men. In order to reduce the impact of other factors on the volunteers' fatigue and to focus on the self-view mode only, the interview questions used in the experiment will all be simple interview questions that will not significantly increase the volunteers' cognitive load. Some examples of interview questions are shown in Table 2. In addition, all experiments will be carried out in a specialist soundproof room laboratory at the department of Information Systems in University of Galway. A portable, flexible, wearable EEG acquisition device will allow the volunteers to focus more on the video conference and will minimise the impact of the EEG acquisition device on the volunteers, so a 14-channel wireless EEG headset was used in this study. The position of each channel follows the International 10–20 Montage System [19], referenced to linked ears and sampled at 256 Hz. The topographic map is shown in Fig. 1 and their names are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. The Volunteers will be unified using a 14-inch MacBook Pro screen.

EEG-Based Fatigue Measurement Framework

Many studies have shown that the power ratio between different frequency bands of EEG can be used to detect human mental fatigue. Fatigue is associated with

Table 2 Some examples of interview questions

Session no	Questions
1	What is your favorite color?
	What is your favorite animal?
	Do you have any pets?
	…
2	What are your hobbies?
	Do you collect anything?
	Who is your favorite superhero?
	…

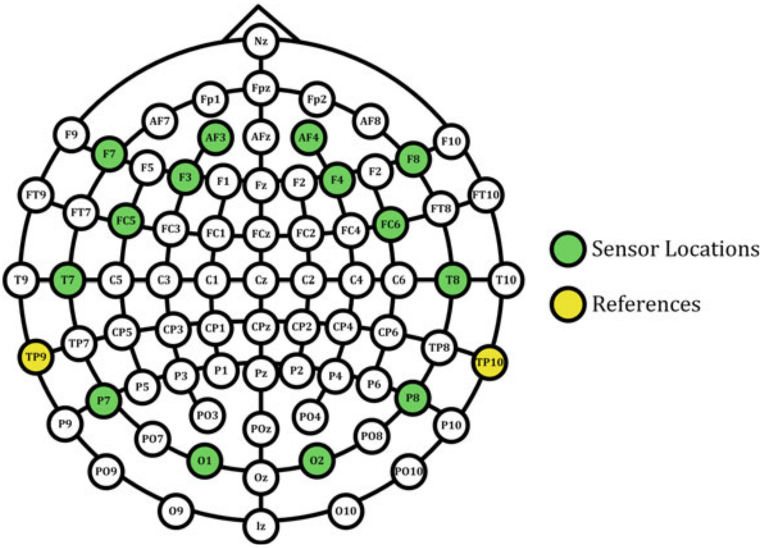


Fig. 1 The topographic map of channel position

significant changes in brainwave activity. The work of Ashley Craig et al. found that spectral activity significantly increased at the EEG θ, α1, and α2 bands when a person is fatigued [18]. In this study, a novel EEG-based fatigue monitoring framework will be provided which will use the slow wave EEG activity as the monitor to observe human mental fatigue. There is still no consensus on the definition of the different EEG frequency bands between different studies. The EEG frequency bands we use are as follows: δ (0.5–3.5 Hz), θ (4–7.5 Hz), α1 (8–10 Hz), α2 (10.5–13 Hz), and β (14–30 Hz). The overview of the framework is shown in Fig. 2 and it has three steps:

1. EEG acquisition, which corresponds to Sect. 4.1 above.
2. EEG cleaning, where a specialist EEGLAB plug-in is used to remove artifacts from sources such as eyes and muscles using ICA and related strategies [20].

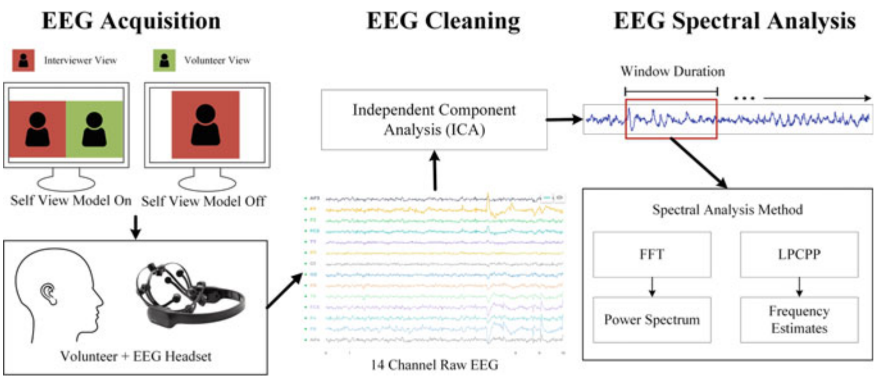


Fig. 2 Overview of the EEG-based fatigue measurement framework

3. EEG spectral analysis, where the EEG signal from a single channel is split into 1 s window size signals and applied spectral analysis method. There are two kinds of spectral analysis methods used here:
   a. One is Fast Fourier Transform (FFT) which is a typical waveform-based spectral analysis method. It can be used to analyse the frequency content of EEG over time and give us the results of how the EEG power spectrum is distributed [11, 12, 14–16, 18].
   b. Another one is a recently proposed parameterised-based spectral analysis method called Linear Predictive Coding Pole Processing (LPCPP). This method for EEG spectral feature extraction and directly gives us numerical estimation frequency results [21–24].

These two typical spectral analysis methods, FFT and LPCPP, will be used to observe the EEG power spectrum and the number of changes in the EEG dominant frequency estimates respectively to see the differences in EEG spectral and therefore to observe the differences in fatigue. The details of the experimental results are in Sect. 4.3.

Experimental Results

Two forms of spectral results from the FFT and LPCPP will be used to analyse the EEG spectral activity, the power spectrum and the dominant frequency estimates. These results were further processed to measure EEG spectral activity. One is Average Spectral Power (ASP) which is used to measure the spectral power changes results. Another one is the Probability Distribution Function (PDF) which is used to describe the probability of EEG dominant frequency estimates. The purpose of this study is to observe the human fatigue difference between the self-view model on/off. A series of spectral results at the self-view mode on/off using ASP and PDF will be demonstrated here, such as:

1. The differences between different genders (i.e. male and female)
2. The differences between different EEG channel locations.
3. The differences between the different EEG bands.

5 Current Progress and Future Direction

Currently, we have recruited 40 volunteers to take part in the study, of which half are male and half are female, all of them from the University of Galway. The data acquisition is expected to be completed by the end of March. We plan to spend 2–3 months cleaning and analysing the EEG data to identify if turning on/off self-view mode is more likely to produce VCF. In this study, an EEG-based framework for VCF analysis is presented. The design of the experiment focuses on the self-view mode, a typical feature of VCF in video conferencing. The analysis of EEG signals can help us to build on our insights by precisely determining how engagement with self-view mode in video conference platforms affects user fatigue levels. The results of these analyses will be used to inform the design of video conferencing platforms and improve the user experience of video conferencing. An output of the experiment will be the creation of a new EEG dataset which involves 40 volunteers. In the future, more kinds of biological signals (e.g. ECG, EMG, EOG) could be considered to help provide more insights for the optimisation of video conferencing platforms.

Acknowledgements This work has emanated from research conducted with the financial support of the Silicon Valley Community Partnership.

References

1. Riedl, R. (2022). On the stress potential of videoconferencing: Definition and root causes of Zoom fatigue. *Electronic Markets, 32*, 153–177.
2. Morris, B. (2020). Why does Zoom exhaust you? Science has an answer. *Wall Street Journal, 27*.
3. Nadler, R. (2020). Understanding “zoom fatigue”: Theorizing spatial dynamics as third. *Computers and Composition.*
4. Fauville, G., Luo, M., Queiroz, A. C. M., Bailenson, J. N., & Hancock, J. (2021). Zoom exhaustion & fatigue scale. *Computers in Human Behavior Reports, 4*, 100119.
5. Massner, C. K. (2021). *The use of video conferencing in higher education* (pp. 75–94). Communication Management.
6. Li, J, Maarten, H. L., & René, R. (2022). Fewer faces displayed simultaneously, less videoconference fatigue in distance learning? An experimental study.
7. Montag, C., Rozgonjuk, D., Riedl, R., & Sindermann, C. (2022). On the associations between videoconference fatigue, burnout and depression including personality associations. *Journal of Affective Disorders Reports*, 100409.

8. Bailenson, N. J. (2021). Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, and Behavior*
9. Fauville, G., Luo, M., Muller Queiroz, A. C., Bailenson, J. N., & Hancock, J. (2021). Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. Available at SSRN 3820035.
10. Åkerstedt, T., Torsvall, L., & Gillberg, M. (1987). Sleepiness in shiftwork. A review with emphasis on continuous monitoring of EEG and EOG. *Chronobiology International, 4*, 129–140.
11. Deng, P., Qiu, X., Tang, Z., Zhang, W., Zhu, L., Ren, H., Zhou, G., & Sheng, R. (2020). Detecting fatigue status of pilots based on deep learning network using EEG signals. *IEEE Transactions on Cognitive and Developmental Systems, 13*, 575–585.
12. Acı, Ç. İ., Murat, K., & Yuriy, M. (2019). Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. *Expert Systems with Applications*, 153–166.
13. Subasi, A. (2005). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications*, 701–711.
14. Lal, S. K., Ashley, C., Peter, B., Les, K., & Hung N. (2003). Development of an algorithm for an EEG-based driver fatigue counter measure. *Journal of Safety Research*, 321–328.
15. Jap, B. T., Sara, L., Peter, F., & Evangelos, B. (2009). Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications*, 2352–2359
16. Simon, M., Eike, A. S., Wilhelm, E. K., Martin, F., Andreas, B., Claus, A., Martin, B., Wolfgang, R., & Michael, S. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clinical Neurophysiology* 1168–1178.
17. Trejo, L. J., Karla, K., Roman, R., Rebekah, L. K., & Leslie, D. M. (2015). *EEG-based estimation and classification of mental fatigue* (pp. 472–589).
18. Craig, A., Yvonne, T., Nirupama, W., & Hung, N. (2012). Regional brain wave activity changes associated with fatigue. *Psychophysiology* 575–582.
19. Klem, G. H., Luders, H. O., Jasper, H. H., & Elgar, C. (1999). The ten-twenty electrode system of the International Federation. In G. Deuschl, & A. Eisen (Eds.), *Recommendation for the practice of clinical Neurophysiology: Guidelines of the international federation of neurophysiology* (pp. 3–6). Electroencephalography and Clinical Neurophysiology.
20. Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 9–21.
21. Jin, X., Mark, D., & Ruairi, D. F. (2020). A robust LPC filtering method for time-resolved morphology of EEG activity analysis. In *26th annual conference of the section of bioengineering of the royal academy of medicine in Ireland*.
22. Jin, X., Mark, D., & Ruairi, D. F. (2020). New robust LPC-based method for time-resolved morphology of high-noise multiple frequency signals. In *31st Irish Signals and Systems Conference (ISSC)* (pp. 1–6).
23. Jin, X., Davis, M., & de Fréin, R. (2021). A linear predictive coding filtering method for the time-resolved morphology of EEG activity. In *32nd Irish Signals and Systems Conference (ISSC)* (pp. 1–61).
24. Jin, X., Mark, D., & Ruairi, D. F. (2021). An LPC pole processing method for enhancing the identification of dominant spectral features. *Electronics Letters*, 708–710.

Fatigue and Stress Levels in Digital Collaboration: A Pilot Study with Video Conferencing and the Metaverse

Fabian J. Stangl, René Riedl, Wolfgang J. Weitzl, and Sebastian Martin

Abstract This pilot study investigated the effects of digital collaboration technologies on heart rate variability (HRV), fatigue, and perceived stress. Experimental data were collected from university students who performed a digital collaboration task in either the metaverse or MS Teams. Heart rate (HR) was measured at baseline and throughout the task using an electrocardiogram-based measurement device (Polar H7 chest strap). HRV data (time domain metrics) and self-reported data were compared during and after the task and between groups. The results show that digital collaboration technologies cause a decrease in parasympathetic activity (RMSSD) with higher self-reported stress levels of individuals collaborating in metaverse compared to those working with the videoconferencing tool MS Teams. These results suggest that digital collaboration technologies are related to variations in parasympathetic nervous system activity and perceived stress, suggesting that monitoring autonomic nervous system activity during digital collaboration needs to be considered to counteract symptoms of fatigue or digital stress.

F. J. Stangl (✉) · R. Riedl · W. J. Weitzl
Digital Business Institute, School of Business and Management, University of Applied Sciences Upper Austria, Steyr, Austria
e-mail: fabian.stangl@fh-ooe.at

R. Riedl
e-mail: rene.riedl@fh-ooe.at

W. J. Weitzl
e-mail: wolfgang.weitzl@fh-ooe.at

R. Riedl
Institute of Business Informatics—Information Engineering, Johannes Kepler University Linz, Linz, Austria

S. Martin
Department of Health, Social and Public Management, University of Applied Sciences Upper Austria, Linz, Austria
e-mail: sebastian.martin@fh-ooe.at

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_9

Keywords Autonomic Nervous System (ANS) Activity · Digital Collaboration · Digital Stress · Experimental Design · Fatigue · Heart Rate (HR) · Heart Rate Variability (HRV) · NeuroIS

1 Introduction

The rise of digitalization has brought about significant changes in the way people collaborate with each other. Traditional methods of collaboration, such as face-to-face meetings and phone calls, have been largely replaced by digital collaboration tools and platforms (e.g., MS Teams and Slack) that allow remote and real-time collaboration via the Internet [1, 2]. Digital tools of the *2D-Internet* like video conferencing systems, complement physical reality and lead to a hybrid world [3]. Indeed, one specific trend based on digitalization that has the potential to create such a hybrid world and supposed to revolutionize different domains is the *metaverse* [4–8]. Following [4, 7–10], we define metaverse as an umbrella term for extended reality or cross reality, which covers the integration of various technologies such as augmented reality, virtual reality, and mixed reality [4, 8, 9]. Metaverse is a new generation of the Internet [10] (i.e., *3D-Internet*), in which users are supposed to experience a sense of virtual presence within a particular location or with other individuals rather than mere spectators [7].

Research is of utmost importance to ensure beneficial interactions between humans and technology in business, management, and other organizational contexts such as education. Thereby, digital collaboration—also referred to as electronic or e-collaboration in the literature—encompasses all computer-based modes that facilitate interaction, communication, and coordination among people across diverse contexts (e.g., tools based on the 2D- or 3D-Internet) [11] and can span various levels of analysis, such as individuals, teams (groups), organizations, and societies [12]. Recent advances and emerging opportunities in digital collaboration, especially via metaverse, illustrate the importance of exploring perceived ease of use to enhance user satisfaction and downstream variables such as task performance [3]. Recent research indicates that developments in digital collaboration can also have a 'dark side.' For example, the widespread and continuous use of video conferencing systems like MS Teams or Zoom can result in stress-related depletion of cognitive and physiological resources due to prolonged and inappropriate use, leading to 'video conferencing fatigue' or 'Zoom fatigue' [13–17]. In addition, the proliferation of digital collaboration tools such as MS Teams or Slack in today's workplace, which are designed to enhance communication and facilitate decision-making [18–20], has also brought the potential to cause stress by contributing to constant interruptions during task performance [21–24] (e.g., computer-mediated communication interruptions [25]). Against the background of recent developments in digital collaboration, there is an urgent need both in practice and science to investigate the impact of alternative forms of digital collaboration on users, their performance, and well-being—specifically regarding symptoms of fatigue and digital stress.

The remainder of this paper is structured as follows. Section 2 presents the related work and research contributions in more detail. Section 3 presents the experimental study. The corresponding subsections provide information on the participants, experimental design and procedure, data processing and analysis of the Heart Rate (HR) data. In Sect. 4, the results of our experimental study are presented, followed by limitations in Sect. 5. Finally, Sect. 6 provides implications and concluding remarks.

This study employed two methodological approaches. First, our research propositions in Sect. 2 are based on a literature review of published theoretical and empirical works in the technostress and Neuro-Information Systems (NeuroIS) literature. Second, in Sect. 3, we present insights from an experimental NeuroIS pilot study that combined self-reported data with neurophysiological measurements to explore the affective and cognitive processes involved in digital collaboration. Note that NeuroIS is an interdisciplinary research field at the nexus of neurophysiology and digital technologies within the Information Systems (IS) discipline that applies neuroscience and neurophysiological tools and methods along with self-reported data to better understand the development, use, and impact of digital technologies, such as collaboration in different digital collaboration technologies [26–33]. As additional substantiation for a NeuroIS approach to support our research, we also followed the call for NeuroIS research by vom Brocke et al. [34], which recognizes the affective and cognitive effects of digital technology use as a topic of societal importance that should be examined through a neuroscientific lens.

2 Related Work and Contribution

The ubiquitous use of digital technologies in the workplace has also substantiated the severity of digital stress or so-called *technostress* [35–38]—a concept that has the potential to affect well-being and health [39], which is especially true in the long run [40]. Digital technologies, particularly smartphones and enterprise social collaboration platforms (e.g., MS Teams, Slack), have contributed considerably to work-related interruptions [21, 22, 25, 41], which can occur, for example, as expected or unexpected interruptions at any time during task execution [42]. This can result in elevated stress levels and adverse effects on performance and productivity (for a review, please see [43]). Despite the numerous advantages of using digital collaboration technologies, such as improved access to information, efficient communication, and productivity gains in organizations [44], the rapid advancement of technology and its increasing global use [45] have led to new challenges that require examination in both scientific research and practice. For example, a recent study found that remote work can lead to a more static and siloed collaboration network among information workers, resulting in fewer connections between the different parts [46]. Another contribution is that remote work can increase stress levels, resulting in a reduction in perceived productivity and subsequent job satisfaction [47]. In addition, excessive use of digital technologies has negative psychological and physiological effects [18, 37, 43, 48]. Hence, developing strategies that promote psychological well-being [3] and facilitate efficient work practices [41, 49], such as preventing fatigue and digital stress during digital collaboration, is critical to counteracting the potential negative consequences of digital technology use [50], which could pose serious issues for both the economy and society as a whole [37, 40, 51].

A review of the technostress literature focusing on digital collaboration reveals empirical studies that address the stress potential of various digital technologies. For example, Chen and Karahanna [24] investigated how interruptions, a major trigger for digital stress [21–23, 52], affect individuals through various technologies. They observed that interruptions resulting from work-related phone calls or instant messages during leisure time can lead to negative outcomes owing to an excessive number of interruptions. On the other hand, email interruptions can have a mixed effect on individuals, as they can provide a sense of accomplishment when a task is completed while easing the psychological transition between work and non-work activities. Note that research has identified various technological characteristics that are relevant in the context of technology and stress. Drawing on Ayyagari et al. [37] and Ragu-Nathan et al. [38], these characteristics can be mainly categorized into four types: (i) usability characteristics (i.e., usefulness, complexity, and reliability), (ii) dynamic characteristics (i.e., pace of change), (iii) intrusive characteristics (i.e., presenteeism, anonymity), and (iv) individual characteristics (e.g., age, gender, education, and self-efficacy in digital technologies). Consequently, when considering interaction via different digital collaboration technologies, we can deductively conclude that such technologies have different potentials to induce stress, especially when considering variations across 2D- or 3D-Internet platforms. Accordingly, we propose the following research proposition (RP) in this study: **Perceived stress in interaction varies between different digital collaboration technologies**.

To gain a deeper understanding of the affective and cognitive processes involved in the use and impact of digital collaboration technologies, we measured HR during digital collaboration. As outlined by the seminal NeuroIS research agenda contribution, HR can serve as an indicator of an individual's physiological state, reflecting alterations in cognitive attention when focusing on a particular situation [26]. Beyond individual-level analysis, neurophysiological measurements can also provide insights into the effects of digital technologies on groups, organizations, and society as a whole [53]. One such effect is *fatigue*, also known as cognitive [54] or mental fatigue [55], which refers to a psychophysiological state resulting from cognitive exertion (e.g., caused by task performance in digital collaboration) and mainly depends on the context, dynamics, form, and period of exertion (e.g., collaboration in different digital collaboration technologies) [56, 57]. Therefore, early detection of fatigue is crucial to prevent a decrease in task performance [54, 58, 59], reduction in goal-directed attention [60], feelings of tiredness, lack of energy, or perceived exhaustion [61]. In addition, alterations in bodily signals, such as HR, can affect brain activity and subsequently affect sensory and cognitive performance [62], making fatigue prevention particularly important for individuals working in critical infrastructures, such as aviation [63]. Additionally, bidirectional communication exists between the central nervous system (specifically the brain) and Autonomic Nervous System (ANS), which regulates HR, highlighting their interdependence [40]. Hence, monitoring bodily signals can aid in comprehending users' affective and cognitive processes, which can clarify why and how certain effects occur during digital collaboration [26, 33].

The measurement of HR as an indicator of ANS activity, along with the related measurement of Heart Rate Variability (HRV), seems to be highly suitable for detecting fatigue and preventing its consequences [56, 64–66]. From an IS perspective, it is crucial to consider alterations in the Sympathetic (SNS) and Parasympathetic Nervous Systems (PNS), as they are responsible for maintaining bodily homeostasis and are part of the ANS [28, 40]. Indeed, physiological states induced by fatigue can trigger physiological responses through SNS and PNS, leading to an increase in HR, faster breathing, muscle tension, sweaty palms, and elevated blood pressure [23]. Consequently, when considering the varying stress potential of interaction in different digital collaboration technologies, we can deductively conclude that digital collaboration technologies have different potentials to induce fatigue, especially when comparing digital collaboration technologies based on 2D- or 3D-Internet. Accordingly, we propose the following RP: **Fatigue in interaction varies between different digital collaboration technologies**.

Given the varying effects of digital technologies on perceived stress, as highlighted in the technostress literature [24], and the evidenced diagnostic value of heart-related measures in detecting fatigue [56, 64–66], the aim of this experimental NeuroIS pilot study was to investigate the following research question: **How does interaction via video conferencing and the metaverse, as different digital collaboration technologies, affect users' stress and fatigue?** Considering the seminal contributions to the NeuroIS research agenda (e.g., [26, 27]), our experimental NeuroIS pilot study specifically aims to investigate how heart-related measurements, as part of ANS activity, can contribute to a deeper understanding of the affective and cognitive processes involved in digital collaboration in different digital collaboration technologies.

3 Methods and Materials

In our pilot experimental study comparing fatigue and stress levels across different digital collaboration technologies, we used a between-subjects experiment with two conditions: digital collaboration (i) via metaverse (i.e., 3D-Internet) and (ii) via MS Teams (i.e., 2D-Internet). Participants were recruited from a pool of undergraduate business students at the University of Applied Sciences Upper Austria who received course credit for their participation. This study was conducted on campus.

The condition assignment was randomized and performed by one of the authors of this study. The randomized controlled trial was conducted on two different days (one day per condition). The SoSci survey platform (www.soscisurvey.de) was used as the self-report stress questionnaire during the experiment. To simulate digital collaboration, each participant performed a digital collaboration task in a separate room at an individual workstation.

Participants

For our experimental pilot study, we recruited four participants aged between 23 and 32 years old (mean [±SD] age: 26.5 ± 3.87 years; median 25.5 years), comprising of 2 men and 2 women, equally divided between the two experimental conditions. The participants were all in good fitness with an overall mean body mass index of 24 ± 3.32 kg/m^2. Standard exclusion criteria, as per [44, 67], were applied during participant recruitment, excluding those who smoked, drank alcohol, took medication, had acute or chronic hormonal dysregulation, or had psychosomatic or psychiatric illnesses. Participants were also instructed to abstain from consuming alcohol or caffeine and engaging in physical activity after 7.00 p.m. the day before the experiment and to consume only water two hours before their participation.

Experimental Design and Procedure

Participants received a letter with instructions (e.g., refraining from smoking) for participation one week before and as a reminder 24 h before the experiments. Upon arrival, the participant was greeted by an instructor and the experimental procedure was explained. Subsequently, the participant was provided with an electrocardiogram-based measurement device (Polar H7 chest strap) to measure HR (for an overview of different methodological approaches to measure HR, please see [68, pp. 8–9]) throughout the experiment and was seated in a comfortable chair to perform the experimental task. The HR signal was displayed on the smartphone to ensure data collection. Note that this measurement device has been previously employed in NeuroIS research (e.g., [69–72]), indicating its suitability for HR measurement. Moreover, validation studies have recommended the use of this measurement device in consumer, clinical, and research settings [73–75], further supporting our choice of measurement device. For the metaverse condition, we used the Meta Quest 2 virtual reality headset (https://www.meta.com/quest/products/quest-2/).

At the beginning of the experiment, participants received a 10-to 15-min introduction to the digital collaboration task environment (either metaverse or MS Teams), which included a comprehensive overview of all the features needed for digital collaboration. In the metaverse condition, participants were provided with an additional ten-minute familiarization period to customize their avatar and become familiar with the Meta Quest 2 touch controller as well as to Horizon Workrooms as the metaverse environment. This approach aims to reduce the possibility of unexpected experiences during digital collaboration, which could potentially affect the results. Following this introduction, the participants watched a standardized 5-min relaxation video for baseline HR measurements. Subsequently, the experimental task involved collaborating on a product launch using designated digital collaboration technology.

During the digital collaboration sessions, an online instructor was available in addition to the two physical instructors to address any issues that may arise during the collaboration process or with technology. After the 30-min marketing-related task, the participants filled out a standardized online questionnaire that included demographic information, such as age, gender, and weight, as well as questions about perceived stress. To measure self-reported stress, we used the 5-item situational stress scale by Hennig-Thurau et al. [76] (5-point Likert scale; sample item: "After the task, I feel exhausted.").

Data Processing and Analysis of HR Data

To process the collected HR data and calculate HRV, we used Kubios Oy's HRV standard analysis software (https://www.kubios.com/hrv-standard/), which is known for its advanced and user-friendly analysis capabilities [77] and has been employed in other scientific studies (e.g., [69]). Note that this software offers various features, including tools for eliminating artifacts such as missed or spurious beats, analysis methods in the time and frequency domains, and the ability to compute less commonly used features such as entropy measures or measures derived from a recurrence plot (for an overview of different applied methods to measure and analyze HRV, see also [77, pp. 270–272]). To further the quality of the HR data, we removed the first and last 30 s of both the baseline and task performance periods as artifacts can significantly impact HR and HRV metrics [30, 79].

For data analysis, we decided to focus on alterations in ANS function by examining alterations in selected HRV metrics in the SNS and PNS compared between the baseline and task performance periods. More specifically, we primarily focused our analysis on the alteration of the following time domain HRV metrics: Bavesky Stress Index and Root Mean Square of Successive Differences (RMSDD), which have also been used as metrics in NeuroIS research [78]. The Bavesky Stress Index reflects cardiovascular system stress, where high Stress Index values indicate lowered HRV and increased SNS activity [80, 81]. RMSSD reflects rapid changes in beat-to-beat intervals in the heart [82]. This parameter is indicative of strong PNS activation, with higher values indicating increased PNS activity [79] (for an overview of HRV metrics, please see [79]; for a recent application of this metric in IS research, see [83]).

4 Results

In this section, we present the main findings of our pilot study in which we compare fatigue and stress levels across two different digital collaboration technologies. Overall, the results suggest that digital collaboration technologies can decrease

PNS activities (RMSSD). Furthermore, the findings imply that self-reported stress is higher in the metaverse condition than in the video conferencing situation.

Alterations in ANS based on HRV—The results indicate that digital collaboration caused a decrease in PNS activity. During the baseline measurement, the mean PNS index of all four participants was −1.21 ± 0.08 and decreased to a mean of −1.34 ± 0.19 during task performance, indicating fatigue during digital collaboration. Specifically, the mean PNS index at baseline measurement was -1.16 ± 0.05 for metaverse and −1.25 ± 0.09 for MS Teams, and at task performance was −1.19 ± 0.12 for metaverse and −1.48 ± 0.09 for MS Teams, suggesting that digital collaboration technology may additionally moderate the effect on fatigue. This finding is supported by the Bavesky Stress Index. During the baseline measurement, the mean Bavesky Stress Index value of all four participants was 22.3 ± 3.84 and decreased to a mean of 17.1 ± 2.51 during task performance, indicating decreased SNS activity during digital collaboration. Specifically, the mean Bavesky Stress Index at baseline measurement was 23.7 ± 5.94 for metaverse and 21.0 ± 1.20 for MS Teams, and at task performance was 18.4 ± 1.98 for metaverse and 15.8 ± 2.9 for MS Teams, suggesting that digital collaboration technology may additionally moderate the effect on SNS activity. Another HRV metric that supported this finding was RMSSD. During the baseline measurement, the mean RMSSD value of the four participants was 8.4 ± 1.24 and increased to a mean of 9.6 ± 1.88 during task performance, indicating increased PNS activity during digital collaboration. Specifically, the mean RMSSD at baseline measurement was 8.60 ± 1.41 for metaverse and 8.20 ± 1.56 for MS Teams, and at task performance was 8.75 ± 0.91 for metaverse and 10.4 ± 2.62 for MS Teams, suggesting that digital collaboration technology may additionally moderate the effect on PNS activity. Table 1 summarizes the results of the alterations in the ANS based on HRV in our experimental NeuroIS pilot study.

Self-Reported Stress—The differences between the metaverse and MS Teams conditions are also mirrored by the results of the self-reported stress levels: Here, the average stress level for the two participants in the metaverse condition was 1.60 ± 0.85 as compared to 1.40 ± 0.28 in the MS Teams condition. However, the overall perceived stress was relatively low in both groups. Table 2 summarizes the results of self-reported stress in our experimental NeuroIS pilot study.

Table 1 Result of alterations in ANS based on HRV

Metric	Period	Full sample	Metaverse	MS teams
Overall PNS Index	Baseline	−1.21 ± 0.08	−1.16 ± 0.05	−1.25 ± 0.09
Overall PNS Index	Task	−1.34 ± 0.19	−1.19 ± 0.12	−1.48 ± 0.09
Bavesky Stress Index	Baseline	22.3 ± 3.84	23.7 ± 5.94	21.0 ± 1.20
Bavesky Stress Index	Task	17.1 ± 2.51	18.4 ± 1.98	15.8 ± 2.90
RMSSD	Baseline	8.4 ± 1.24	8.60 ± 1.41	8.20 ± 1.56
RMSSD	Task	9.6 ± 1.88	8.75 ± 0.91	10.4 ± 2.62

Table 2 Result of self-reported stress

Metric	Full sample	Metaverse	MS teams
Self-reported stress	1.50 ± 0.53	1.60 ± 0.85	1.40 ± 0.28

5 Limitations

While interpreting our preliminary results, it is important to acknowledge the limitations of this study. First, the sample size used in this experimental pilot study was small, which may have affected the generalizability and, hence, limited the conclusiveness of our findings. Nonetheless, we intend to replicate this study to validate our exploratory results. Second, our study accounted for common confounders of heart-related measures by excluding participants who smoked, drank alcohol, took medication, had acute or chronic hormonal dysregulation, or had psychosomatic or psychiatric illnesses. However, since HR can be influenced by various factors, such as autonomic, respiratory, circulatory, endocrine, and mechanical influences of the heart [84], there may still be confounding variables that we were unable to control for. For example, respiration in hypertensive patients can affect the variance of the heartbeat and therefore must also be considered [85]. Third, it is important to note that we conducted a between-subjects study with two conditions, as the experimental session lasted approximately 70 to 80 min, and we wanted to avoid any potential carry-over effects that may arise from collaborating on the launch of a similar product. However, a within-subject design could be considered to improve the group-level reliability of different tasks (e.g., [86]), although it may not be practical with regard to fatigue measurement. Another approach, in this context, may be to investigate more specific factors. For example, a recent study investigated human–machine interaction by manipulating the subjective perception of risk through a virtual reality environment simulating work at high altitudes [87]. The comparison of different environments, such as working in an office versus a safety–critical workplace, could also be considered a factor in a mixed-factorial design. Overall, future research addressing these limitations could provide further insight into the affective and cognitive processes involved in digital collaboration with various digital collaboration technologies.

6 Implications and Concluding Remarks

Advances in technology have revolutionized the way we work and collaborate with others, allowing us more opportunities for digital collaboration than ever before. With the rise of remote work, video conferencing, and virtual collaboration platforms we are no longer limited by physical distance when it comes to working together on projects or sharing ideas [1, 2]. Nevertheless, digital technology presents potential risks to both mental and physical well-being. The specific objective of our experimental pilot study (which is part of a larger ongoing experimental study) was to explore how heart-related measurements, as components of ANS activity, can enhance our understanding of the affective and cognitive processes underlying digital collaboration. Specifically, as the first NeuroIS research study that investigated a potential link between heart-related measurements, fatigue, and perceived stress [88], we found preliminary indications that digital collaboration technology may influence not only stress levels, but also fatigue. While we acknowledge that the sample size of our experimental pilot study is limited, and hence definitive conclusions are not appropriate, it is worth noting that future research with a larger sample should validate this finding on digital collaboration, which would benefit both the economy and society [37, 40, 51].

Acknowledgements This research was funded by the Austrian Science Fund (FWF) as part of the project "Technostress in Organizations" (project number: P 30865), by the Federal State of Upper Austria through the Austrian Research Promotion Agency (FFG) as part of the project "Interruption", and by the Federal State of Upper Austria as part of the project "Irrational Digital Behaviors" at the University of Applied Sciences Upper Austria.

References

1. Davis, A., Murphy, J., Owens, D., Khazanchi, D., & Zigurs, I. (2009). Avatars, people, and virtual worlds: Foundations for research in metaverses. *Journal of the Association for Information Systems, 10*(2), 90–117. https://doi.org/10.17705/1jais.00183
2. Owens, D., Mitchell, A., Khazanchi, D., ZIgurs, I. (2011). An empirical investigation of virtual world projects and metaverse technology capabilities. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 42*(1), 74–101. https://doi.org/10.1145/1952712.1952717
3. Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., Guo, Y., Harris, D., Ioannou, A., Jeong, K. (Kate), Konomi, S., Krömker, H., Kurosu, M., Lewis, J. R., Marcus, A., Meiselwitz, G., Moallem, A., Mori, H., Nah, F.F.-H., Ntoa, S., Rau, P.-L. P., Schmorrow, D., Siau, K., Streitz, N., Wang, W., Yamamoto, S., Zaphiris, P., Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction, 35*(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259
4. Mystakidis, S. (2022). Metaverse. *Encyclopedia, 2*(1), 486–497. https://doi.org/10.3390/encyclopedia2010031
5. Buhalis, D., Lin, M. S., & Leung, D. (2023). Metaverse as a driver for customer experience and value co-creation: Implications for hospitality and tourism management and marketing. *International Journal of Contemporary Hospitality Management, 35*(2), 701–716. https://doi.org/10.1108/IJCHM-05-2022-0631
6. Dincelli, E., & Yayla, A. (2022). Immersive virtual reality in the age of the Metaverse: A hybrid-narrative review based on the technology affordance perspective. *The Journal of Strategic Information Systems, 31*(2), 101717. https://doi.org/10.1016/j.jsis.2022.101717
7. Peukert, C., Weinhardt, C., Hinz, O., & van der Aalst, W. M. P. (2022). Metaverse: How to approach its challenges from a BISE perspective. *Business & Information Systems Engineering, 64*(4), 401–406. https://doi.org/10.1007/s12599-022-00765-9
8. Xi, N., Chen, J., Gama, F., Riar, M., & Hamari, J. (2022). The challenges of entering the metaverse: An experiment on the effect of extended reality on workload. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10244-x

9. Milgram, P., Takemura, H., Utsumi, A., Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. In: Das, H. (Ed.), *Telemanipulator and Telepresence Technologies: 31 October-1 November 1994, Boston, Massachusetts* (Vol. 2351, pp. 282–292). The International Society for Optical Engineering. https://doi.org/10.1117/12.197321
10. Zuckerberg, M. (2021). Meta: Founder's Letter. https://about.fb.com/news/2021/10/founders-letter/, last accessed 2023/03/13.
11. Riemer, K., Steinfield, C., & Vogel, D. (2009). eCollaboration: On the nature and emergence of communication and collaboration technologies. *Electronic Markets, 19*(4), 181–188. https://doi.org/10.1007/s12525-009-0023-1
12. Gallivan, M. J., & Benunan-Fich, R. (2005). A framework for analyzing levels of analysis issues in studies of e-collaboration. *IEEE Transactions on Professional Communication, 48*(1), 87–104. https://doi.org/10.1109/TPC.2005.843301
13. Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*(1). https://doi.org/10.1037/tmb0000030
14. Riedl, R. (2022). On the stress potential of videoconferencing: Definition and root causes of Zoom fatigue. *Electronic Markets, 32*(1), 153–177. https://doi.org/10.1007/s12525-021-00501-3
15. Montag, C., Rozgonjuk, D., Riedl, R., & Sindermann, C. (2022). On the associations between videoconference fatigue, burnout and depression including personality associations. *Journal of Affective Disorders Reports, 10*, 100409. https://doi.org/10.1016/j.jadr.2022.100409
16. Li, J., Lamers, M. H., & Riedl, R. (2022). Fewer faces displayed simultaneously, less videoconference fatigue in distance learning? An experimental study. In: *Proceedings of the 43rd International Conference on Information Systems.*
17. Riedl, R., Kostoglou, K., Wriessnegger, S. C., & Müller-Putz, G. R. (2023). Videoconference fatigue from a neurophysiological perspective: Experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). *Scientific Reports, 13*(1), 18371. https://doi.org/10.1038/s41598-023-45374-y
18. Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems, 16*(1), 1–47. https://doi.org/10.17705/1jais.00387
19. Jia, L., Huang, L., Yan, Z., Hall, D., Song, J., & Paradice, D. (2019). The importance of policy to effective IM use and improved performance. *Information Technology & People, 33*(1), 180–197. https://doi.org/10.1108/ITP-09-2018-0409
20. McMurtry, K. (2014). Managing email overload in the workplace. *Performance Improvement, 53*(7), 31–37. https://doi.org/10.1002/pfi.21424
21. Stangl, F. J., & Riedl, R. (2023). Interruption science as a research field: Towards a taxonomy of interruptions as a foundation for the field. *Frontiers in Psychology, 14*, 1043426. https://doi.org/10.3389/fpsyg.2023.1043426
22. Stangl, F. J., & Riedl, R. (2023). Interruptions in the workplace: An exploratory study among digital business professionals. In: Nah, F. F.-H., & Siau, K. (Eds.), *HCI in Business, Government and Organizations: 10th International Conference, HCIBGO 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part II.* LNCS (Vol. 14039, pp. 400–422). Springer. https://doi.org/10.1007/978-3-031-36049-7_29
23. Puranik, H., Koopman, J., & Vough, H. C. (2020). Pardon the interruption: An integrative review and future research agenda for research on work interruptions. *Journal of Management, 46*(6), 806–842. https://doi.org/10.1177/0149206319887428
24. Chen, A. J., & Karahanna, E. (2018). Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS Quarterly, 42*(4), 1023–1042. https://doi.org/10.25300/MISQ/2018/13631
25. Russell, E., Jackson, T., & Banks, A. (2021). Classifying computer-mediated communication (CMC) interruptions at work using control as a key delineator. *Behaviour & Information Technology, 40*(2), 191–205. https://doi.org/10.1080/0144929X.2019.1683606

26. Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P.H., Müller-Putz, G.R., Pavlou, P. A., Straub, D. W., vom Brocke, J., Weber, B. (2010). On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. *Communications of the Association for Information Systems, 27*(15), 243–264. https://doi.org/10.17705/1CAIS.02715
27. Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G. R., Benbasat, I., Gefen, D., Kenning, P. H., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702. https://doi.org/10.2307/41703475
28. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain*. Springer. https://doi.org/10.1007/978-3-662-45091-8
29. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 51*(3), 13–54. https://doi.org/10.1145/3410977.3410980
30. Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems, 15*(10), I–XXXV. https://doi.org/10.17705/1jais.00377
31. Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research, 22*(4), 687–702. https://doi.org/10.1287/isre.1100.0284
32. Loos, P., Riedl, R., Müller-Putz, G. R., vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering, 2*(6), 395–401. https://doi.org/10.1007/s12599-010-0130-8
33. Riedl, R., Davis, F. D., Banker, R. D., & Kenning, P. H. (2017). *Neuroscience in information systems research: Applying knowledge of brain functionality without neuroscience tools*. Springer. https://doi.org/10.1007/978-3-319-48755-7
34. vom Brocke, J., Hevner, A., Léger, P.-M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*(1), 9–24. https://doi.org/10.1080/0960085X.2019.1708218
35. Brod, C. (1984). *Technostress: The human cost of the computer revolution*. Addison-Wesley Publishing.
36. Weil, M. M., & Rosen, L. D. (1997). *TechnoStress: Coping with technology @work @home @play*. Wiley.
37. Ayyagari, R., Grover, V., & Purvis, R. (2011). Technostress: Technological antecedents and implications. *MIS Quarterly, 35*(4), 831–858. https://doi.org/10.2307/41409963
38. Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433. https://doi.org/10.1287/isre.1070.0165
39. Pflügner, K., Maier, C., Mattke, J., & Weitzel, T. (2021). Personality profiles that put users at risk of perceiving technostress: A qualitative comparative analysis with the Big Five personality traits. *Business & Information Systems Engineering, 63*(4), 389–402. https://doi.org/10.1007/s12599-020-00668-7
40. Riedl, R. (2013). On the biology of technostress: Literature review and research agenda. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 44*(1), 18–55. https://doi.org/10.1145/2436239.2436242
41. Tams, S., Ahuja, M., Thatcher, J. B., & Grover, V. (2020). Worker stress in the age of mobile technology: The combined effects of perceived interruption overload and worker control. *Journal of Strategic Information Systems, 29*(1), 101595. https://doi.org/10.1016/j.jsis.2020.101595
42. Ortiz de Guinea, A., & Webster, J. (2013). An investigation of information systems use patterns: Technological events as triggers, the effect of time, and consequences for performance. *MIS Quarterly, 37*(4), 1165–1188. https://doi.org/10.25300/MISQ/2013/37.4.08

43. Fischer, T., & Riedl, R. (2017). Technostress research: A nurturing ground for measurement pluralism? *Communications of the Association for Information Systems, 40*(17), 375–401. https://doi.org/10.17705/1CAIS.04017
44. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2012). Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering, 4*(2), 61–69. https://doi.org/10.1007/s12599-012-0207-7
45. Couffe, C., & Michael, G. A. (2017). Failures due to interruptions or distractions: A review and a new framework. *The American Journal of Psychology, 130*(2), 163–181. https://doi.org/10.5406/amerjpsyc.130.2.0163
46. Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., & Teevan, J. (2021). The effects of remote work on collaboration among information workers. *Nature Human Behaviour, 6*(1), 43–54. https://doi.org/10.1038/s41562-021-01196-4
47. Toscano, F., & Zappalà, S. (2020). Social isolation and stress as predictors of productivity perception and remote work satisfaction during the COVID-19 pandemic: The role of concern about the virus in a moderated double mediation. *Sustainability, 12*(23), 9804. https://doi.org/10.3390/su12239804
48. Stangl, F. J., Riedl, R., Kiemeswenger, R., & Montag, C. (2023). Negative psychological and physiological effects of social networking site use: The example of Facebook. *Frontiers in Psychology, 14*, 1141663. https://doi.org/10.3389/fpsyg.2023.1141663
49. Baham, C., Kalgotra, P., Nasirpouri Shadbad, F., Sharda, R. (2022). Generational differences in handling technology interruptions: A qualitative study. *European Journal of Information Systems,* 1–21. https://doi.org/10.1080/0960085X.2022.2070557
50. Tarafdar, M., Gupta, A., & Turel, O. (2013). The dark side of information technology use. *Information Systems Journal, 23*(3), 269–275. https://doi.org/10.1111/isj.12015
51. Tarafdar, M., Tu, Q., & Ragu-Nathan, T. S. (2010). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems, 27*(3), 303–334. https://doi.org/10.2753/MIS0742-1222270311
52. Stangl, F. J., & Riedl, R. (2023). Digital Detoxing als Maßnahme gegen digitalen Stress: Auf die Verwendung digitaler Technologien temporär verzichten. *Theologisch-praktische Quartalschrift, 171*(3), 274–283.
53. Riedl, R. (2009). Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen. *NeuroPsychoEconomics, 4*(1), 32–44.
54. Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In: Ackerman, P. L. (Ed.), *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications* (pp. 167–187). American Psychological Association. https://doi.org/10.1037/12343-008
55. Darnai, G., Matuz, A., Alhour, H. A., Perlaki, G., Orsi, G., Arató, Á., Szente, A., Áfra, E., Nagy, S. A., Janszky, J., & Csathó, Á. (2023). The neural correlates of mental fatigue and reward processing: A task-based fMRI study. *NeuroImage, 265*, 119812. https://doi.org/10.1016/j.neuroimage.2022.119812
56. Matuz, A., van der Linden, D., Darnai, G., & Csathó, Á. (2022). Generalisable machine learning models trained on heart rate variability data to predict mental fatigue. *Scientific Reports, 12*(1), 20023. https://doi.org/10.1038/s41598-022-24415-y
57. Phillips, R. O. (2015). A review of definitions of fatigue – And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour, 29*, 48–56. https://doi.org/10.1016/j.trf.2015.01.003
58. Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45*(1–3), 73–93. https://doi.org/10.1016/S0301-0511(96)05223-4
59. Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews, 59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

60. Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research, 25*(1), 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011
61. Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: Focus on definition and measurement. *Sleep Medicine Reviews, 10*(1), 63–76. https://doi.org/10.1016/j.smrv.2005.05.004
62. Park, H.-D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience, 17*(4), 612–618. https://doi.org/10.1038/nn.3671
63. Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003
64. Gao, R., Yan, H., Duan, J., Gao, Y., Cao, C., Li, L., & Guo, L. (2022). Study on the nonfatigue and fatigue states of orchard workers based on electrocardiogram signal analysis. *Scientific Reports, 12*(1), 4858. https://doi.org/10.1038/s41598-022-08705-z
65. Melo, H. M., Nascimento, L. M., & Takase, E. (2017). Mental fatigue and heart rate variability (HRV): The time-on-task effect. *Psychology & Neuroscience, 10*(4), 428–436. https://doi.org/10.1037/pne0000110
66. Laurent, F., Valderrama, M., Besserve, M., Guillard, M., Lachaux, J.-P., Martinerie, J., & Florence, G. (2013). Multimodal information improves the rapid detection of mental fatigue. *Biomedical Signal Processing and Control, 8*(4), 400–408. https://doi.org/10.1016/j.bspc.2013.01.007
67. Takahashi, T. (2005). Social memory, social stress, and economic behaviors. *Brain Research Bulletin, 67*(5), 398–402. https://doi.org/10.1016/j.brainresbull.2005.06.006
68. Stangl, F. J., Riedl, R. (2022). Measurement of heart rate and heart rate variability with wearable devices: A systematic review. In: *Proceedings of the 17th International Conference on Wirtschaftsinformatik*.
69. Baumgartner, D., Fischer, T., Riedl, R., & Dreiseitl, S. (2019). Analysis of heart rate variability (HRV) feature robustness for measuring technostress. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., & Randolph, A. B. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2018*. LNISO (Vol. 29, pp. 221–228). Springer. https://doi.org/10.1007/978-3-030-01087-4_27
70. Fischer, T., & Riedl, R. (2020). Technostress measurement in the field: A case report. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Fischer, T. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2020*. LNISO (Vol. 43, pp. 71–78). Springer. https://doi.org/10.1007/978-3-030-60073-0_9
71. Jahn, K., Kordyaka, B., Ressing, C., Roeding, K., & Niehaves, B. (2020). Designing self-presence in immersive virtual reality to improve cognitive performance—A research proposal. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Fischer, T. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019*. LNISO (Vol. 32, pp. 83–91). Springer. https://doi.org/10.1007/978-3-030-28144-1_9
72. Kalischko, T., Fischer, T., Riedl, R. (2020). Techno-unreliability: A pilot study in the field. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Fischer, T. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019*. LNISO (Vol. 32, pp. 137–145). Springer. https://doi.org/10.1007/978-3-030-28144-1_15
73. Etiwy, M., Akhrass, Z., Gillinov, L., Alashi, A., Wang, R., Blackburn, G., Gillinov, S. M., Phelan, D., Gillinov, A. M., Houghtaling, P. L., Javadikasgari, H., & Desai, M. Y. (2019). Accuracy of wearable heart rate monitors in cardiac rehabilitation. *Cardiovascular Diagnosis and Therapy, 9*(3), 262–271. https://doi.org/10.21037/cdt.2019.04.08
74. Pasadyn, S. R., Soudan, M., Gillinov, M., Houghtaling, P., Phelan, D., Gillinov, N., Bittel, B., Desai, M. Y. (2019). Accuracy of commercially available heart rate monitors in athletes: A prospective study. *Cardiovascular Diagnosis and Therapy, 9*(4), 379–385. https://doi.org/10.21037/cdt.2019.06.05

75. Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology, 2*(1), 104. https://doi.org/10.1001/jamacardio.2016.3340
76. Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., & Kübler, R. V. (2022). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science*. https://doi.org/10.1007/s11747-022-00908-0
77. Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine, 113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024
78. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A review of NeuroIS research with a focus on applied methods. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., &Müller-Putz, G. R. (eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 269–283). Springer. https://doi.org/10.1007/978-3-031-13064-9_28
79. Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health, 5*, 258. https://doi.org/10.3389/fpubh.2017.00258
80. Baevsky, R. M. (2002). Analysis of heart rate variability in space medicine. *Human Physiology, 28*(2), 202–213.
81. Baevsky, R. M., Chernikova, A. G. (2017) Heart rate variability analysis: Physiological foundations and main methods. *Cardiometry,* (10), 66–76. https://doi.org/10.12710/cardiometry.2017.10.6676
82. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology (1996): Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation, 93*(5), 1043–1065. https://doi.org/10.1161/01.CIR.93.5.1043
83. Nadj, M., Rissler, R., Adam, M. T. P., Knierim, M. T., Li, M. X., Mädche, A., & Riedl, R. (2023). What disrupts flow in office work? A NeuroIS study on the impact of frequency and relevance of IT-mediated interruptions. *MIS Quarterly, 47*(4), 1615–1646. https://doi.org/10.25300/MISQ/2023/17200
84. Quintana, D. S., Alvares, G. A., & Heathers, J. A. J. (2016). Guidelines for reporting articles on psychiatry and heart rate variability (GRAPH): Recommendations to advance research communication. *Translational Psychiatry, 6*(5), e803. https://doi.org/10.1038/tp.2016.73
85. Ruediger, H., Seibt, R., Scheuch, K., Krause, M., & Alam, S. (2004). Sympathetic and parasympathetic activation in heart rate variability in male hypertensive patients under mental stress. *Journal of Human Hypertension, 18*(5), 307–315. https://doi.org/10.1038/sj.jhh.1001671
86. Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A. B. M., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., & Meyer-Lindenberg, A. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *NeuroImage, 60*(3), 1746–1758. https://doi.org/10.1016/j.neuroimage.2012.01.129
87. Hoesterey, S., & Onnasch, L. (2022). A new experimental paradigm to manipulate risk in human-automation research. *Human Factors*. In press.https://doi.org/10.1177/00187208221133878
88. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability in NeuroIS research: Review of empirical results. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Müller-Putz, G. R. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 285–299). Springer. https://doi.org/10.1007/978-3-031-13064-9_29

Neurophysiological Measurements in the Research Field of Digital Detoxing: Review and Implications for Future Research

Fabian J. Stangl and René Riedl

Abstract Over the past decade, stress caused by the ubiquity of digital technologies has increasingly come into focus in both scientific research and practice. One strategy that has received increasing attention in recent years to counteract the negative consequences of digital stress is digital detoxing, which refers to temporary or complete disengagement from digital technologies. To lay the foundation for future research, we reviewed existing empirical studies on digital detoxing from a neurophysiological measurement perspective. The identified empirical digital detoxing studies that included neurophysiological measurements are described along with the following factors: research objective, research method, sample size, study population, and research findings. We conclude that this increasingly relevant research topic is still at a relatively nascent stage, both in general and specifically regarding neurophysiological measurements. We anticipate that the use of measures related to the brain and human body, including nervous system activity and hormone measures, will gain relevance in future empirical research on digital detoxing.

Keywords Brain · Digital Detoxing · Digital Detox Research · Digital Stress · Neurophysiological Measures · NeuroIS · Literature Review

F. J. Stangl (✉) · R. Riedl
Digital Business Institute, School of Business and Management, University of Applied Sciences Upper Austria, Steyr, Austria
e-mail: fabian.stangl@fh-steyr.at

R. Riedl
e-mail: rene.riedl@fh-steyr.at

R. Riedl
Institute of Business Informatics—Information Engineering, Johannes Kepler University Linz, Linz, Austria

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_10

1 Introduction

The impact of digital technologies is pervasive in today's society and economy and influences our daily lives. For example, due to the lockdowns caused by COVID-19 and the need for social distancing, companies, universities, and other organizations had to respond quickly and implement web-conferencing systems (e.g., Microsoft Teams or Zoom) so that employees, students, and other people could go about their daily tasks and activities [1]. In general, it is a well-established fact that the use of digital technologies may come along with significant benefits, including increased access to information, rapid communication possibilities, and productivity gains in organizations [2]. However, digital technologies also increase in complexity and require research to promote beneficial interactions between humans and technology [3]. Over the past decade, stress and fatigue caused by the use and ubiquity of digital technologies has received increasing attention in both scientific research and practice. This research has also provided ample evidence that digital stress, also referred to as technostress in the literature, along with related phenomena such as videoconference fatigue [4–7], is a serious problem for both society and the economy [6, 8, 9]. In addition to negative effects on task performance and productivity [10–12], digital stress can lead to various negative psychological and physiological effects, including higher psychological or physiological stress levels or reduced physical or psychological well-being [6, 11, 13, 14]. Hence, research activities and findings from research on digital stress are particularly valuable to counteract this "dark side" of digital technologies [15].

The proliferation of technology and its global spread, along with the acceleration of daily life, are striking modern developments [16]. In an increasingly digital world, interruptions during task execution, for example, are one of the major triggers of digital stress that individuals must contend with [17–20]. Interruptions are indeed a ubiquitous aspect of the workplace, with system breakdowns [2, 21] and variable response times [22–26] being just a few examples. While individual interruptions may not seem to take much time, such as an incoming instant messenger message or email [27], the sheer volume of interruptions that occur during a workday can be problematic. For example, empirical research has found that notifications from digital devices and programs used in the workplace interrupt employees four to six times per working hour [10]. The time required to cope with these interruptions adds up, posing a challenge to modern workplaces. Empirical studies indicate that interruptions can result in a loss of work time, ranging from 5% [28] to approximately 28% [29]. These exemplary research findings on interruptions as triggers of digital stress highlight the potential for the increasing use of digital technologies in the workplace to cause stress, ultimately impacting the well-being and health of employees [30], particularly in the long term [8].

One strategy that has received increasing attention in both scientific research and practice to counteract the negative effects of the use of digital technologies and the accompanying acceleration of processes in work and life is *digital detoxing*. Drawing on prior definitions [31, 32] and research [33, 34], digital detoxing can be defined as temporary or complete disengagement from digital technologies (e.g., temporary abstinence from social media such as Facebook, Instagram, and/or Snapchat), while also considering strategies to reduce exposure to digital technologies (e.g., regular breaks from computer work). In this regard, research has indicated that temporary abstinence or reduction from digital technology use may indeed be a promising strategy to mitigate the negative psychological and physiological effects of digital technology use. For example, Brailovskaia et al. [35] found a significant decrease in perceived depressive symptoms and an increase in perceived life satisfaction over a 14-day period when participants reduced their daily use of Facebook to 20 min. Other studies have also shown that a temporary break from Facebook can notably boost perceived life satisfaction [36] and reduce cortisol levels as a physiological stress indicator [37], thus supporting the above finding. Hence, research on digital detoxing has the potential to improve overall well-being and promote balance and equanimity in digital technology use [31].

Neurophysiological measurements can thereby offer valuable insights into the underlying neural mechanisms associated with digital detoxing and its effects, helping to deepen our understanding of how digital technologies impact human cognition, emotions, and behavior, and their associated consequences. For example, Walden et al. [38] investigated whether and how students' brain structures changed over the course of a semester. Using diffusion tensor imaging, they found that the connectivity between regions associated with intelligence generally increased. Such findings could also be helpful in understanding the effects of digital detoxing (e.g., alteration of cognitive functions or alteration of physiological indicators) and in drawing more definitive conclusions about the effects (e.g., [39]). In essence, neurophysiological measurements can provide objective measures of the effectiveness of digital detoxing interventions, which can also increase the validity of research findings [40]. Assessing the impact of digital technologies is crucial, as they pose not only psychological risks, such as reduced perceived life satisfaction, but also physiological risks. For example, research has explored the association between Facebook use and human brain alterations [41–43]. He et al. [41] found that excessive social media use is associated with decreased gray matter volume in the bilateral amygdala and right ventral striatum, indicating differences in key neural systems in individuals with high behavioral patterns of excessive social media use. Hence, neurophysiological measurements in digital detoxing can provide valuable information on how digital technologies affect the brain and body, leading to a better understanding of their consequences and potential benefits at both psychological and physiological levels.

To advance digital detoxing from a neurophysiological measurement perspective, we examined empirical literature on digital detoxing (i.e., [36, 37, 44–106]) to provide insights into the appropriate use of various neurophysiological measurement methods [107, 108]. Note that this paper is based on the results of a recent systematic literature review, which reviewed the methodological aspects of existing empirical digital detoxing studies [33]. To the best of our knowledge, this recent literature review is also the most comprehensive review of the currently available peer-reviewed empirical digital detoxing studies. However, the current paper goes beyond this original review by examining empirical studies on the application of neurophysiological measurements. Thus, the objective is to lay a foundation for future studies by analyzing previous research from a neurophysiological measurement perspective. Based on the results of our original analysis of 65 studies published in peer-reviewed journals and conference proceedings [33], in the present paper we address the following research question: **What neurophysiological measurements have been applied to examine digital detoxing?**

The remainder of this paper is structured as follows. Section 2 describes the methodology used in this review. The results are presented in Sect. 3. Section 4 discusses possible future research avenues for advancing digital detoxing research. Finally, in Sect. 5, we provide concluding remarks.

2 Review Methodology

The starting point of our literature search was a publication by Mirbabaie, Stieglitz, and Marx [31], which includes the first conceptualization of digital detoxing in the Information Systems discipline. For the subsequent literature search, we considered peer-reviewed journals and conference publications in the following eight databases: ACM Digital Library, AIS eLibrary, Emerald Group Publishing, EBSCO Information Services, IEEE Xplore, Science Direct, Scopus, and Web of Science. For the literature search, we used generic terms that represent digital detoxing (e.g., "digital abstinence" [93] or "digital disconnection" [86]) and specific terms used as characteristics in the different research fields (e.g., "digital free tourism" [105] or "offline tourism" [92]). The review process was based on existing recommendations for conducting literature searches [109, 110]. Further details on the review methodology can be found in [33].

Overall, our review methodology ensures broad and valid consideration for empirical digital detoxing studies. Specifically, our literature base consisted of 65 empirical digital detoxing studies published before and in June 2022. Note that other systematic literature reviews on digital detoxing have identified a much smaller number of studies despite the inclusion of non-peer-reviewed papers. For example, Özdemir and Goktas [111] identified 20 studies, and Radtke et al. [112] identified 21 studies. Based on the analysis of the full text of the 65 studies, we could only identify three papers that included neurophysiological measurements in their empirical digital detoxing study (i.e., [37, 46, 76]).

3 Review Results

This section presents the main findings of the literature review. Specifically, to lay a foundation for future research activities on digital detoxing research, we describe the three identified studies in more detail by describing the following factors: research objective, research method, sample size, study population, and research outcomes.

Measuring the Effects of Digital Detoxing with Actigraphy

Dunican et al. [46] conducted a longitudinal quasi-experimental study investigating the impact of the evening use of electronic devices on sleep quality and next-day athletic and cognitive performance. Although this study does not have a direct focus on neurophysiological measurements, we included it in our review because it focuses on body-related parameters that are related to neurophysiology (e.g., the stress hormone cortisol affects sleep; [2]). Eighteen elite judo athletes (1 female, mean age 18 $\pm$ 2 years between 16 and 24) wore a wearable device for six consecutive days and nights to obtain continuous actigraphy measurements of sleep quantity and quality, while nine participants assigned themselves to the device-restricted experimental group and removed all electronic devices for a 48-h period. The researchers found that all athletes significantly overestimated their sleep duration and underestimated their time of sleep onset when comparing subjective and actigraphy-based measures of sleep. However, no differences in physical (with a single leg three hop test) or cognitive function (with Cogstate Research software) were observed between the groups.

Measuring the Effects of Digital Detoxing with Electrodermal Activity

Anrijs et al. [76] conducted a longitudinal quasi-experimental study to investigate whether digital detoxing is effective in reducing perceived short-term stress and whether this could be objectively measured using markers of both smartphone usage and physiological stress. During two consecutive weeks, fifteen students (10 female; mean age 22 years between 21 and 24) wore a wearable device that measured physiological stress based on electrodermal activity (also referred to as galvanic skin response or skin conductance; [113, 114]) and installed an application on their smartphones to track which applications were used and for how long. The first week was a regular week of smartphone use, and the second week was a digital detox week. Based on the analysis of 10 students, the study found that participants used fewer applications during the digital detox week than during the regular week, and their electrodermal activity levels were significantly lower, indicating lower stress levels.

Measuring the Effects of Digital Detoxing with Salivary Cortisol

Vanman et al. [37] conducted a longitudinal experimental study with a randomized controlled trial design (RCT) to investigate the impact of Facebook abstinence on perceived stress and subjective well-being. One hundred and thirty-eight active Facebook users (87 female; mean age 22.43 years between 18 and 40) who were randomly assigned to either a condition where they were asked to abstain from using Facebook for five days (i.e., experimental condition) or continue using it as usual (i.e., control condition). Perceived stress (with the 10-item Perceived Stress Scale [115]) and subjective well-being (with the 5-item Satisfaction with Life Scale [116]) were assessed before and after the test period, along with the measurement of salivary cortisol. The results showed that participants in the experimental condition had lower levels of cortisol, but also experienced a decrease in life satisfaction compared to those in the control condition.

4 Research Potentials

We contribute to this research by providing a neurophysiological perspective on digital detoxing. Overall, our review shows that several neurophysiological measures were used to measure digital detoxing effects, including actigraphy [46], electrodermal activity [76], and salivary cortisol [37]. Table 1 summarizes important method factors, namely research objective, research method, neurophysiological measurement, sample size, study population, and major research outcomes.

Building on the results of our systematic literature review, we derived three major potential avenues for future research. *First*, conducting studies with neurophysiological measurements can be considered a promising approach to advance digital detoxing research. For example, Anrijs et al. [76] showed that digital detoxing can significantly reduce physiological stress even in a short time by measuring electrodermal activity. Another example is Vanman et al. [37], who found lower cortisol levels when abstaining from Facebook for five days using salivary cortisol as a hormone measure to examine the effects of digital detoxing. For the NeuroIS researchers, the selection of appropriate tools for measuring the effects of digital detoxing, however, depends on the research context [40], which can be categorized into three types: measurement of the Central Nervous System (CNS), measurement of the Peripheral Nervous System (PNS), which includes the subsystem of the Autonomic (ANS) and Somatic Nervous Systems (SNS), and measurement of the hormone system [114]. The accuracy of the various tools available for these three categories can be evaluated in terms of temporal resolution, which describes the time between stimulus onset and measurement of the physiological signal. Moreover, tools used for measuring the CNS can also be evaluated based on their spatial resolution, which refers to the accuracy in pinpointing the location of brain activity (for an overview of

Table 1 Neurophysiological measurements in reviewed digital detoxing studies

Study Detail	Dunican et al. [46]	Anrijs et al. [76]	Vanman et al. [37]
Research Objective	Effect on sleep quality, athletic and cognitive performance	Effect on perceived short-term stress	Effect on perceived stress and subjective well-being
Research Method	Longitudinal quasi-experimental study	Longitudinal quasi-experimental study	Longitudinal experimental study with RCT design
Neurophysiological Measurement	Actigraphy	Electrodermal activity	Salivary cortisol
Sample Size	18 (1 female, mean age 18 ± 2 years between 16 and 24)	15 (10 female; mean age 22 years between 21 and 24); only 10 for analysis	138 (87 female; mean age 22.43 years between 18 and 40)
Study Population	Elite judo athletes	Students	Active Facebook users
Research Outcomes	No differences in physical or cognitive function, overestimation of sleep duration, underestimation of sleep duration	Lower electrodermal activity and less smartphone application use during digital detoxing	Lower levels of salivary cortisol and decrease in life satisfaction when abstaining from Facebook use

neurophysiological tools with a discussion of the strengths and weaknesses of each measurement method per research setting, please see [115, pp. 47–72]; for a detailed discussion of methods used in cognitive neuroscience, please see [117]). Overall, as demonstrated by studies measuring the effects of digital detoxing through ANS measurements with electrodermal activity [76] and hormone system measurements with cortisol levels [37], neurophysiological measurements hold great potential for future research on digital detoxing.

Second, technological advancements in wearable devices have facilitated the monitoring and tracking of physiological indicators, such as actigraphy measurements of sleep quantity and quality [46] or measurements of electrodermal activity for physiological activation and stress [76]. Wearable devices are becoming increasingly relevant in the toolbox of NeuroIS researchers. One main advantage of wearable devices is their low degree of intrusiveness, which refers to the extent to which a measurement instrument interferes with an ongoing task [40]. Wearables are minimally intrusive because they allow for a high degree of movement freedom during (experimental) task execution (e.g., temporary disengagement from social media), can be used in natural positions while sitting or walking, and are non-invasive because they do not require insertion into the body or attachment of electrodes to the body or scalp. Consequently, wearables are ideal tools for conducting human–computer interaction (and also NeuroIS) studies in both laboratory and field environments [40]. Prior to conducting empirical studies, it is crucial to carefully evaluate and consider general quality criteria for measurement methods in psychometrics and psychophysiology [40], such as the reliability and validity of wearable devices [118], along with ethical, legal, and societal implications [119, 120]. For example, Dunican et al. [46] used the Fatigue Science Readiband™ wearable device to measure sleep quality and quantity, which can be recommended for sleep studies in healthy adults based on validation studies for consumer, clinical, and research purposes [121, 122]. Overall, the increasing technological progress in wearable devices, combined with their low degree of intrusiveness, high degree of movement freedom, and ability to be used in both laboratory and field environments, represent a promising avenue for future NeuroIS research.

Third, empirical studies aiming to extend the current digital detoxing research findings using neurophysiological measurements could explore additional measures that could indicate alterations in cognitive functions or physiological markers. Various neurophysiological measurements can provide valuable information in different research contexts related to digital detoxing. For example, electroencephalography can be used to examine changes in brainwave activity in response to digital detox interventions [123], whereas functional magnetic resonance imaging can help identify changes in brain activity associated with cognitive or emotional processing during and after detoxing [124]. For example, He et al. [41] used structural magnetic resonance imaging to investigate the association between excessive social media use and gray matter volume in key neural systems. In addition to these brain imaging methods, another research approach is to use physiological indicators as an index of ANS activity to extend current digital detoxing research. For example, Heart Rate (HR) and Heart Rate Variability (HRV) as indices of ANS function [125] could provide valuable information for different measurement purposes, including physiological arousal or physiological stress during task performance [126]. Considering technological progress, such as different methods for measuring HR and HRV [127], the use of ANS activity indicators could provide physiological insights into the effects of digital detoxing at the individual level, which would also facilitate the development of personalized digital detoxing intervention strategies. For example, depending on individual indicators, employees can be advised to take regular breaks from computer work to reduce exposure to digital technologies, ultimately increasing the effectiveness of digital detoxing and digital technology use in general. Notably, empirical research on breaks in demanding computer work using cardiovascular, electrodermal, and electromyographic measurements has shown that short breaks (7.5 min after 50 min of work) are more effective in promoting recovery from mental and emotional strain until the early afternoon, while a longer break (15 min after 100 min of work) is more effective in reducing fatigue and emotional strain in the late afternoon [128]. Importantly, neurophysiological measurements are not limited to the individual level, as they can also be used to examine the effects on the group, organization, or societal levels [129]. Overall, investigating digital detoxing and its effectiveness with measures that have not yet been applied in this research field, such as brain imaging methods or ANS activity measurement with cardiovascular measures, holds great future research potential.

5 Concluding Remarks

Most current research on digital detoxing relies on surveys, content analysis, or experiments [33], with only a small fraction of experiments taking neurophysiological measures into consideration. Indeed, we found only three papers in empirical research on digital detoxing that included neurophysiological measures (i.e., [37, 46, 76]). Given these findings, we argue that research on this increasingly relevant topic is still in a relatively nascent stage, particularly from a measurement perspective.

Future empirical studies should consider the neurophysiological measurements. Importantly, these measures should be used as complements to self-reports (e.g., psychometric measures; [37]) or other methods (e.g., informal dialogues; [92]), not as substitutes [130]. For example, Vanman et al. [37] investigated the effect of digital detoxing using self-report measures of perceived stress and subjective well-being, and by measuring salivary cortisol. Moreover, neurophysiological measures can also advance our understanding of the effects of digital technologies on the brain (e.g., [41]) and support the development of effective digital detoxing intervention strategies to reduce the negative effects of long exposure to digital technologies (e.g., regular breaks from computer work to decrease sympathetic activity and increase parasympathetic activity). Overall, we anticipate the increasing use of neurophysiological measurements in digital detox research.

Acknowledgements This research was funded by the Austrian Science Fund (FWF) as part of the project "Technostress in Organizations" (project number: P 30865) and by the Federal State of Upper Austria through the Austrian Research Promotion Agency (FFG) as part of the project "Interruption" at the University of Applied Sciences Upper Austria.

References

1. Hacker, J., vom Brocke, J., Handali, J., Otto, M., & Schneider, J. (2020). Virtually in this together – How web-conferencing systems enabled a new virtual togetherness during the COVID-19 crisis. *European Journal of Information Systems, 29*(5), 563–584. https://doi.org/10.1080/0960085X.2020.1814680
2. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2012). Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering, 4*(2), 61–69. https://doi.org/10.1007/s12599-012-0207-7
3. Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., Guo, Y., Harris, D., Ioannou, A., Jeong, K. (Kate), Konomi, S., Krömker, H., Kurosu, M., Lewis, J. R., Marcus, A., Meiselwitz, G., Moallem, A., Mori, H., Nah, F. F.-H., Ntoa, S., Rau, P.-L. P., Schmorrow, D., Siau, K., Streitz, N., Wang, W., Yamamoto, S., Zaphiris, P., Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction, 35*(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259
4. Riedl, R. (2022). On the stress potential of videoconferencing: Definition and root causes of Zoom fatigue. *Electronic Markets*, *32*(1), 153–177. https://doi.org/10.1007/s12525-021-00501-3

5. Riedl, R., Kostoglou, K., Wriessnegger, S. C., & Müller-Putz, G. R. (2023). Videoconference fatigue from a neurophysiological perspective: Experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). *Scientific Reports, 13*(1), 18371. https://doi.org/10.1038/s41598-023-45374-y
6. Ayyagari, R., Grover, V., & Purvis, R. (2011). Technostress: Technological antecedents and implications. *MIS Quarterly, 35*(4), 831–858. https://doi.org/10.2307/41409963
7. Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433. https://doi.org/10.1287/isre.1070.0165
8. Riedl, R. (2013). On the biology of technostress: Literature review and research agenda. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 44*(1), 18–55. https://doi.org/10.1145/2436239.2436242
9. Tarafdar, M., Tu, Q., & Ragu-Nathan, T. S. (2010). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems, 27*(3), 303–334. https://doi.org/10.2753/MIS0742-1222270311
10. Mirhoseini, S., Hassanein, K., Head, M., Watter, S. (2020). User performance in the face of IT interruptions: The role of executive functions. In: Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., and Fischer, T. (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019.* LNISO (Vol. 32, pp. 41–51). Springer. https://doi.org/10.1007/978-3-030-28144-1_5
11. Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems, 16*(1), 1–47. https://doi.org/10.17705/1jais.00387
12. Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes, 109*(2), 168–181. https://doi.org/10.1016/j.obhdp.2009.04.002
13. Fischer, T., & Riedl, R. (2017). Technostress research: A nurturing ground for measurement pluralism? *Communications of the Association for Information Systems, 40*(17), 375–401. https://doi.org/10.17705/1CAIS.04017
14. Stangl, F. J., Riedl, R., Kiemeswenger, R., & Montag, C. (2023). Negative psychological and physiological effects of social networking site use: The example of Facebook. *Frontiers in Psychology, 14*, 1141663. https://doi.org/10.3389/fpsyg.2023.1141663
15. Tarafdar, M., Gupta, A., & Turel, O. (2013). The dark side of information technology use. *Information Systems Journal, 23*(3), 269–275. https://doi.org/10.1111/isj.12015
16. Couffe, C., & Michael, G. A. (2017). Failures due to interruptions or distractions: A review and a new framework. *The American Journal of Psychology, 130*(2), 163–181. https://doi.org/10.5406/amerjpsyc.130.2.0163
17. Chen, A. J., & Karahanna, E. (2018). Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS Quarterly, 42*(4), 1023–1042. https://doi.org/10.25300/MISQ/2018/13631
18. Puranik, H., Koopman, J., & Vough, H. C. (2020). Pardon the interruption: An integrative review and future research agenda for research on work interruptions. *Journal of Management, 46*(6), 806–842. https://doi.org/10.1177/0149206319887428
19. Stangl, F. J., & Riedl, R. (2023). Interruption science as a research field: Towards a taxonomy of interruptions as a foundation for the field. *Frontiers in Psychology, 14*, 1043426. https://doi.org/10.3389/fpsyg.2023.1043426
20. Stangl, F. J., & Riedl, R. (2023). Interruptions in the workplace: An exploratory study among digital business professionals. In: F. F.-H. Nah & K. Siau (Eds.), *HCI in Business, Government and Organizations: 10th International Conference, HCIBGO 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part II.* LNCS (Vol. 14039, pp. 400–422). Springer. https://doi.org/10.1007/978-3-031-36049-7_29
21. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2013). Computer breakdown as a stress factor during task completion under time pressure: Identifying gender differences based on

skin conductance. *Advances in Human-Computer Interaction, 2013*, 420169. https://doi.org/10.1155/2013/420169
22. Kohlisch, O., & Kuhmann, W. (1997). System response time and readiness for task execution the optimum duration of inter-task delays. *Ergonomics, 40*(3), 265–280. https://doi.org/10.1080/001401397188143
23. Kuhmann, W., Boucsein, W., Schäfer, F., & Alexander, J. (1987). Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction. *Ergonomics, 30*(6), 933–943. https://doi.org/10.1080/00140138708969789
24. Riedl, R., Fischer, T. (2018). System response time as a stressor in a digital world: Literature review and theoretical model. In: F. F.-H. Nah & B. S. Xiao (Eds.), *HCI in Business, Government, and Organizations: 5th International Conference, HCIBGO 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings.* LNISA (Vol. 10923, pp. 175–186). Springer. https://doi.org/10.1007/978-3-319-91716-0_14
25. Kohrs, C., Hrabal, D., Angenstein, N., & Brechmann, A. (2014). Delayed system response times affect immediate physiology and the dynamics of subsequent button press behavior. *Psychophysiology, 51*(11), 1178–1184. https://doi.org/10.1111/psyp.12253
26. Schäfer, F. (1990). The effect of system response times on temporal predictability of work flow in human-computer interaction. *Human Performance, 3*(3), 173–186. https://doi.org/10.1207/s15327043hup0303_3
27. Russell, E., Jackson, T., & Banks, A. (2021). Classifying computer-mediated communication (CMC) interruptions at work using control as a key delineator. *Behaviour & Information Technology, 40*(2), 191–205. https://doi.org/10.1080/0144929X.2019.1683606
28. Gupta, A., & Sharda, R. (2008). SIMONE: A simulator for interruptions and message overload in network environments. *International Journal of Simulation and Process Modelling, 4*(3/4), 237. https://doi.org/10.1504/IJSPM.2008.023685
29. Chen, A. J., & Karahanna, E. (2014). Boundaryless technology: Understanding the effects of technology-mediated interruptions across the boundaries between work and personal life. *AIS Transactions on Human-Computer Interaction, 6*(2), 16–36. https://doi.org/10.17705/1thci.00059
30. Pflügner, K., Maier, C., Mattke, J., & Weitzel, T. (2021). Personality profiles that put users at risk of perceiving technostress: A qualitative comparative analysis with the Big Five personality traits. *Business & Information Systems Engineering, 63*(4), 389–402. https://doi.org/10.1007/s12599-020-00668-7
31. Mirbabaie, M., Stieglitz, S., & Marx, J. (2022). Digital detox. *Business & Information Systems Engineering, 64*(2), 239–246. https://doi.org/10.1007/s12599-022-00747-x
32. Syvertsen, T., & Enli, G. (2020). Digital detox: Media resistance and the promise of authenticity. *Convergence: The International Journal of Research into New Media Technologies, 26*(5–6), 1269–1283. https://doi.org/10.17705/1thci.00059
33. Hager, N., Stangl, F. J., & Riedl, R. (2023). Digital detox research: An analysis of applied methods and implications for future studies. In: *Proceedings of the 18th International Conference on Wirtschaftsinformatik*.
34. Stangl, F. J., & Riedl, R. (2023). Digital Detoxing als Maßnahme gegen digitalen Stress: Auf die Verwendung digitaler Technologien temporär verzichten. *Theologisch-praktische Quartalschrift, 171*(3), 274–283.
35. Brailovskaia, J., Ströse, F., Schillack, H., & Margraf, J. (2020). Less Facebook use – More well-being and a healthier lifestyle? An experimental intervention study. *Computers in Human Behavior, 108*, 106332. https://doi.org/10.1016/j.chb.2020.106332
36. Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior, and Social Networking, 19*(11), 661–666. https://doi.org/10.1089/cyber.2016.0259
37. Vanman, E. J., Baker, R., & Tobin, S. J. (2018). The burden of online friends: The effects of giving up Facebook on stress and well-being. *The Journal of Social Psychology, 158*(4), 496–508. https://doi.org/10.1080/00224545.2018.1453467

38. Walden, E. A., Browne, G. J., & Oboyle, M. (2015). Computational thinking: Changes to the human connectome associated with learning to program. In: *Proceedings of the 36th International Conference on Information Systems.*
39. Léger, P.-M., Sénecal, S., Courtemanche, F., Ortiz de Guinea, A., Titah, R., Fredette, M., & Labonté-LeMoyne, É. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research. *Journal of the Association for Information Systems, 15*(10), 651–678. https://doi.org/10.17705/1jais.00376
40. Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems, 15*(10), I–XXXV. https://doi.org/10.17705/1jais.00377
41. He, Q., Turel, O., Brevers, D., & Bechara, A. (2017). Excess social media use in normal populations is associated with amygdala-striatal but not with prefrontal morphology. *Psychiatry Research: Neuroimaging, 269*, 31–35. https://doi.org/10.1016/j.pscychresns.2017.09.003
42. He, Q., Turel, O., & Bechara, A. (2018). Association of excessive social media use with abnormal white matter integrity of the corpus callosum. *Psychiatry Research: Neuroimaging, 278*, 42–47. https://doi.org/10.1016/j.pscychresns.2018.06.008
43. Montag, C., Markowetz, A., Blaszkiewicz, K., Andone, I., Lachmann, B., Sariyska, R., Trendafilov, B., Eibes, M., Kolb, J., Reuter, M., Weber, B., & Markett, S. (2017). Facebook usage on smartphones and gray matter volume of the nucleus accumbens. *Behavioural Brain Research, 329*, 221–228. https://doi.org/10.1016/j.bbr.2017.04.035
44. Díaz-Meneses, G., & Estupinán-Ojeda, M. (2022). The outbreak of digital detox motives and their public health implications for holiday destinations. *International Journal of Environmental Research and Public Health, 19*(3), 1548. https://doi.org/10.3390/ijerph19031548
45. Dickinson, J. E., Hibbert, J. F., & Filimonau, V. (2016). Mobile technology and the tourist experience: (Dis)connection at the campsite. *Tourism Management, 57*, 193–201. https://doi.org/10.1016/j.tourman.2016.06.005
46. Dunican, I. C., Martin, D. T., Halson, S. L., Reale, R. J., Dawson, B. T., Caldwell, J. A., Jones, M. J., & Eastwood, P. R. (2017). The effects of the removal of electronic devices for 48 hours on sleep in elite judo athletes. *Journal of Strength and Conditioning Research, 31*(10), 2832–2839. https://doi.org/10.1519/JSC.0000000000001697
47. Egger, I., Lei, S. I., & Wassler, P. (2020). Digital free tourism – An exploratory study of tourist motivations. *Tourism Management, 79*, 104098. https://doi.org/10.1016/j.tourman.2020.104098
48. Eide, T. A., Aarestad, S. H., Andreassen, C. S., Bilder, R. M., & Pallesen, S. (2018). Smartphone restriction and its effect on subjective withdrawal related scores. *Frontiers in Psychology, 9*, 1444. https://doi.org/10.3389/fpsyg.2018.01444
49. El-Khoury, J., Haidar, R., Kanj, R. R., Bou Ali, L., & Majari, G. (2021). Characteristics of social media "detoxification" in university students. *Libyan Journal of Medicine, 16*(1), 1846861. https://doi.org/10.1080/19932820.2020.1846861
50. Fan, D. X. F., Buhalis, D., & Lin, B. (2019). A tourist typology of online and face-to-face social contact: Destination immersion and tourism encapsulation/decapsulation. *Annals of Tourism Research, 78*, 102757. https://doi.org/10.1016/j.annals.2019.102757
51. Fioravanti, G., Prostamo, A., & Casale, S. (2020). Taking a short break from Instagram: The effects on subjective well-being. *Cyberpsychology, Behavior, and Social Networking, 23*(2), 107–112. https://doi.org/10.1089/cyber.2019.0400
52. Floros, C., Cai, W., McKenna, B., & Ajeeb, D. (2021). Imagine being off-the-grid: Millennials' perceptions of digital-free travel. *Journal of Sustainable Tourism, 29*(5), 751–766. https://doi.org/10.1080/09669582.2019.1675676
53. Fryman, S., & Romine, W. (2021). Measuring smartphone dependency and exploration of consequences and comorbidities. *Computers in Human Behavior Reports, 4*, 100108. https://doi.org/10.1016/j.chbr.2021.100108
54. Amato, A., Rovai, M., & Andreoli, M. (2019). Digital detox tourism as a resource for the enhancement of cultural heritage. A development study for Certosa di Pisa in Calci (Tuscany).

In: *Proceedings of the International Forum on World Heritage and Legacy Culture Creativity Contamination.*
55. Hall, J. A., Xing, C., Ross, E. M., & Johnson, R. M. (2021). Experimentally manipulating social media abstinence: Results of a four-week diary study. *Media Psychology, 24*(2), 259–275. https://doi.org/10.1080/15213269.2019.1688171
56. Handa, M., & Ahuja, P. (2020). Disconnect to detox: A study of smartphone addiction among young adults in India. *Young Consumers, 21*(3), 273–287. https://doi.org/10.1108/YC-12-2019-1077
57. Hanley, S. M., Watt, S. E., & Coventry, W. (2019). Taking a break: The effect of taking a vacation from Facebook and Instagram on subjective well-being. *PLoS ONE, 14*(6), e0217743. https://doi.org/10.1371/journal.pone.0217743
58. Hassan, T. H., Salem, A. E., & Saleh, M. I. (2022). Digital-free tourism holiday as a new approach for tourism well-being: Tourists' attributional approach. *International Journal of Environmental Research and Public Health, 19*(10), 5974. https://doi.org/10.3390/ijerph19105974
59. Helsper, E. J., & Reisdorf, B. C. (2013). A quantitative examination of explanations for reasons for internet nonuse. *Cyberpsychology, Behavior, and Social Networking, 16*(2), 94–99. https://doi.org/10.1089/cyber.2012.0257
60. Hinsch, C., & Sheldon, K. M. (2013). The impact of frequent social Internet consumption: Increased procrastination and lower life satisfaction. *Journal of Consumer Behaviour, 12*(6), 496–505. https://doi.org/10.1002/cb.1453
61. Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology, 37*(10), 751–768. https://doi.org/10.1521/jscp.2018.37.10.751
62. Jiang, Y., & Balaji, M. S. (2022). Getting unwired: What drives travellers to take a digital detox holiday? *Tourism Recreation Research, 47*(5–6), 453–469. https://doi.org/10.1080/02508281.2021.1889801
63. Jorge, A. (2019). Social media, interrupted: Users recounting temporary disconnection on Instagram. *Social Media + Society, 5*(4), 1–19. https://doi.org/10.1177/2056305119881691
64. Karlsen, F. (2020). Digital detox camp: Values and motivations for engaging in digital disconnect. In: *Proceedings of the 21st Annual Conference of the Association of Internet Researchers.* https://doi.org/10.5210/spir.v2020i0.11244
65. Anashkina, N., Shmatko, M., & Tkachenko, O. (2020). Digital-detox technology for vocational guidance selection of applicants for media communications area. *Mediaobrazovanie, 60*(4), 570–580.
66. Karlsen, F., & Ytre-Arne, B. (2022). Intrusive media and knowledge work: How knowledge workers negotiate digital media norms in the pursuit of focused work. *Information, Communication & Society, 25*(15), 2174–2189. https://doi.org/10.1080/1369118X.2021.1933561
67. Keller, J., Roitzheim, C., Radtke, T., Schenkel, K., & Schwarzer, R. (2021). A mobile intervention for self-efficacious and goal-directed smartphone use in the general population: Randomized controlled trial. *JMIR mHealth and uHealth, 9*(11), e26397. https://doi.org/10.2196/26397
68. Kent, R. (2020). Self-tracking health over time: From the use of Instagram to perform optimal health to the protective shield of the digital detox. *Social Media + Society, 6*(3), 1–14. https://doi.org/10.1177/2056305120940694
69. Ko, M., Yang, S., Lee, J., Heizmann, C., Jeong, J., Lee, U., Shin, D., Yatani, K., Song, J., & Chung, K.-M. (2015). NUGU: A group-based intervention app for improving self-regulation of limiting smartphone use. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1235–1245). ACM, New York, NY, USA. https://doi.org/10.1145/2675133.2675244
70. Li, J., Pearce, P. L., & Oktadiana, H. (2020). Can digital-free tourism build character strengths? *Annals of Tourism Research, 85*, 103037. https://doi.org/10.1016/j.annals.2020.103037

71. Liao, M., & Sundar, S. S. (2022). #facebookdown: Time to panic or detox? Understanding users' reactions to social media outage. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–8). https://doi.org/10.1145/3491101.3519674
72. Liu, H., Liu, W., Yoganathan, V., & Osburg, V.-S. (2021). COVID-19 information overload and generation Z's social media discontinuance intention during the pandemic lockdown. *Technological Forecasting and Social Change, 166*, 120600. https://doi.org/10.1016/j.techfore.2021.120600
73. Matthes, J., Karsay, K., Hirsch, M., Stevic, A., & Schmuck, D. (2022). Reflective smartphone disengagement: Conceptualization, measurement, and validation. *Computers in Human Behavior, 128*, 107078. https://doi.org/10.1016/j.chb.2021.107078
74. Mirbabaie, M., Marx, J., Braun, L.-M., & Stieglitz, S. (2020). Digital detox – Mitigating digital overuse in times of remote work and social isolation. In: *Proceedings of the 31st Australasian Conference on Information Systems.*
75. Muench, C., Feulner, L., Muench, R., & Carolus, A. (2020). Time to log off: An analysis of factors influencing the willingness to participate in a long-term 'Digital Detox' with the smartphone. In: C. Stephanidis & M. Antona (Eds.), *HCI International 2020 - Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part III.* CCIS (Vol. 1226, pp. 209–216). Springer. https://doi.org/10.1007/978-3-030-50732-9_28
76. Anrijs, S., Bombeke, K., Durnez, W., Van Damme, K., Vanhaelewyn, B., Conradie, P., Smets, E., Cornelis, J., De Raedt, W., Ponnet, K., & De Marez, L. (2018). MobileDNA: Relating physiological stress measurements to smartphone usage to assess the effect of a digital detox. In: C. Stephanidis (Ed.), *HCI International 2018 – Posters' Extended Abstracts: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II.* CCIS (Vol. 851, pp. 356–363). Springer. https://doi.org/10.1007/978-3-319-92279-9_48
77. Mutsvairo, B., Ragnedda, M., & Mabvundwi, K. (2022). 'Our old pastor thinks the mobile phone is a source of evil.' Capturing contested and conflicting insights on digital wellbeing and digital detoxing in an age of rapid mobile connectivity. *Media International Australia,* 1–15. https://doi.org/10.1177/1329878X221090992
78. Nguyen, V. T. (2022). The perceptions of social media users of digital detox apps considering personality traits. *Education and Information Technologies, 27*(7), 9293–9316. https://doi.org/10.1007/s10639-022-11022-7
79. Olson, J. A., Sandra, D. A., Chmoulevitch, D., Raz, A., & Veissière, S. P. L. (2022). A nudge-based intervention to reduce problematic smartphone use: Randomised controlled trial. *International Journal of Mental Health and Addiction.* https://doi.org/10.1007/s11469-022-00826-w
80. Paris, C.M., Berger, E.A., Rubin, S., & Casson, M. (2015). Disconnected and unplugged: Experiences of technology induced anxieties and tensions while traveling. In: I. Tussyadiah & A. Inversini (Eds.), *Information and Communication Technologies in Tourism 2015: Proceedings of the International Conference in Lugano, Switzerland, February 3–6, 2015* (pp. 803–816). Springer. https://doi.org/10.1007/978-3-319-14343-9_58
81. Pawłowska-Legwand, A., & Matoga, Ł. (2021). Disconnect from the digital world to reconnect with the real life: An analysis of the potential for development of unplugged tourism on the example of Poland. *Tourism Planning & Development, 18*(6), 649–672. https://doi.org/10.1080/21568316.2020.1842487
82. Pearce, P. L., & Gretzel, U. (2012). Tourism in technology dead zones: Documenting experiential dimensions. *International Journal of Tourism Sciences, 12*(2), 1–20. https://doi.org/10.1080/15980634.2012.11434656
83. Phillips, B., & Shipps, B. (2022). Problematic technology use: The impact of personality and continued use. *The Journal of the Southern Association for Information Systems, 9*(1), 38–63. https://doi.org/10.17705/3JSIS.00021
84. Reinecke, L., Aufenanger, S., Beutel, M. E., Dreier, M., Quiring, O., Stark, B., Wölfling, K., & Müller, K. W. (2017). Digital stress over the life span: The effects of communication

load and internet multitasking on perceived stress and psychological health impairments in a German probability sample. *Media Psychology, 20*(1), 90–115. https://doi.org/10.1080/15213269.2015.1121832
85. Schmitt, J. B., Breuer, J., & Wulf, T. (2021). From cognitive overload to digital detox: Psychological implications of telework during the COVID-19 pandemic. *Computers in Human Behavior, 124*, 106899. https://doi.org/10.1016/j.chb.2021.106899
86. Schwarzenegger, C., & Lohmeier, C. (2021). Creating opportunities for temporary disconnection: How tourism professionals provide alternatives to being permanently online. *Convergence: The International Journal of Research into New Media Technologies, 27*(6), 1631–1647. https://doi.org/10.1177/13548565211033385
87. Basu, R. (2019). Impact of digital detox on individual performance of the employees. *International Journal of Research and Analytical Reviews, 6*(2), 378–381.
88. Sheldon, K. M., Abad, N., & Hinsch, C. (2011). A two-process view of Facebook use and relatedness need-satisfaction: Disconnection drives use, and connection rewards it. *Journal of Personality and Social Psychology, 100*(4), 766–775. https://doi.org/10.1037/a0022407
89. Skierkowski, D., & Wood, R. M. (2012). To text or not to text? The importance of text messaging among college-aged youth. *Computers in Human Behavior, 28*(2), 744–756. https://doi.org/10.1016/j.chb.2011.11.023
90. Stieger, S., & Lewetz, D. (2018). A week without using social media: Results from an ecological momentary intervention study using smartphones. *Cyberpsychology, Behavior, and Social Networking, 21*(10), 618–624. https://doi.org/10.1089/cyber.2018.0070
91. Sutton, T. (2020). Digital harm and addiction: An anthropological view. *Anthropology Today, 36*(1), 17–22. https://doi.org/10.1111/1467-8322.12553
92. Syvertsen, T. (2022). Offline tourism: Digital and screen ambivalence in Norwegian mountain huts with no internet access. *Scandinavian Journal of Hospitality and Tourism, 22*(3), 195–209. https://doi.org/10.1080/15022250.2022.2070540
93. Turel, O., Cavagnaro, D. R., & Meshi, D. (2018). Short abstinence from online social networking sites reduces perceived stress, especially in excessive users. *Psychiatry Research, 270*, 947–953. https://doi.org/10.1016/j.psychres.2018.11.017
94. Turel, O., & Cavagnaro, D. R. (2019). Effect of abstinence from social media on time perception: Differences between low-and at-risk for social media "addiction" groups. *Psychiatric Quarterly, 90*(1), 217–227. https://doi.org/10.1007/s11126-018-9614-3
95. Ugur, N. G., & Koc, T. (2015). Time for digital detox: Misuse of mobile technology and phubbing. *Procedia - Social and Behavioral Sciences, 195*, 1022–1031. https://doi.org/10.1016/j.sbspro.2015.06.491
96. Umasankar, M., Boopathy, S., Padmavathy, S., Fukey, L. N., & Singh, R. (2022). Disconnect to reconnect: Employee wellbeing through digital detoxing. *Journal of Positive School Psychology, 6*(2), 4463–4673.
97. Vally, Z., & D'Souza, C. G. (2019). Abstinence from social media use, subjective well-being, stress, and loneliness. *Perspectives in Psychiatric Care, 55*(4), 752–759. https://doi.org/10.1111/ppc.12431
98. Baumer, E. P. S., Guha, S., Quan, E., Mimno, D., & Gay, G. K. (2015). Missing photos, suffering withdrawal, or finding freedom? How experiences of social media non-use influence the likelihood of reversion. *Social Media + Society, 1*(2), 1–14. https://doi.org/10.1177/2056305115614851
99. van Wezel, M. M. C., Abrahamse, E. L., & Vanden Abeele, M. M. P. (2021). Does a 7-day restriction on the use of social media improve cognitive functioning and emotional well-being? Results from a randomized controlled trial. *Addictive Behaviors Reports, 14*, 100365. https://doi.org/10.1016/j.abrep.2021.100365
100. Widdicks, K., Remy, C., Bates, O., Friday, A., & Hazas, M. (2022). Escaping unsustainable digital interactions: Toward "more meaningful" and "moderate" online experiences. *International Journal of Human-Computer Studies, 165*, 102853. https://doi.org/10.1016/j.ijhcs.2022.102853

101. Wilcockson, T. D. W., Osborne, A. M., & Ellis, D. A. (2019). Digital detox: The effect of smartphone abstinence on mood, anxiety, and craving. *Addictive Behaviors, 99*, 106013. https://doi.org/10.1016/j.addbeh.2019.06.002
102. Wood, N. T., & Muñoz, C. (2021). Unplugged: Digital detox enhances student learning. *Marketing Education Review, 31*(1), 14–25. https://doi.org/10.1080/10528008.2020.1836973
103. Zhang, M., & Zhang, X. (2022). Between escape and return: Rethinking daily life and travel in selective unplugging. *Tourism Management, 91*, 104521. https://doi.org/10.1016/j.tourman.2022.104521
104. Brown, L., & Kuss, D. J. (2020). Fear of missing out, mental wellbeing, and social connectedness: A seven-day social media abstinence trial. *International Journal of Environmental Research and Public Health, 17*(12), 4566. https://doi.org/10.3390/ijerph17124566
105. Cai, W., & McKenna, B. (2023). Power and resistance: Digital-free tourism in a connected world. *Journal of Travel Research, 62*(2), 290–304. https://doi.org/10.1177/00472875211061208
106. Cesareo, M., Tagliabue, M., Oppo, A., & Moderato, P. (2021). The ubiquity of social reinforcement: A nudging exploratory study to reduce the overuse of smartphones in social contexts. *Cogent Psychology, 8*(1), 1880304. https://doi.org/10.1080/23311908.2021.1880304
107. Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems, 10*(2), 75–105. https://doi.org/10.1080/07421222.1993.11518001
108. Chong, S. W., & Reinders, H. (2021). A methodological review of qualitative research syntheses in CALL: The state-of-the-art. *System, 103*, 102646. https://doi.org/10.1016/j.system.2021.102646
109. Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.
110. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: S. Newell, E. A. Whitley, N. Pouloudi, J. Wareham, & L. Mathiassen (Eds.), *Proceedings of the 17th European Conference on Information Systems* (pp. 2206–2217).
111. Özdemir, M. A., & Goktas, L. S. (2021). Research trends on digital detox holidays: A bibliometric analysis, 2012–2020. *Tourism & Management Studies, 17*(3), 21–35. https://doi.org/10.18089/tms.2021.170302
112. Radtke, T., Apel, T., Schenkel, K., Keller, J., & von Lindern, E. (2022). Digital detox: An effective solution in the smartphone era? A systematic literature review. *Mobile Media & Communication, 10*(2), 190–215. https://doi.org/10.1177/20501579211028647
113. Léger, P. M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior, 34*, 273–283. https://doi.org/10.1016/j.chb.2014.02.011
114. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain*. Springer, Heidelberg. https://doi.org/10.1007/978-3-662-45091-8
115. Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396. https://doi.org/10.2307/2136404
116. Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
117. Senior, C., Russell, T., & Gazzaniga, M. S. (2009). *Methods in mind*. MIT Press.
118. Stangl, F.J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability with wearable devices: A systematic review. In: *Proceedings of the 17th International Conference on Wirtschaftsinformatik*.
119. Dagum, P., & Montag, C. (2019). Ethical considerations of digital phenotyping from the perspective of a healthcare practitioner. In: H. Baumeister & C. Montag (Eds.), *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics*. SNPBE (pp. 13–28). Springer. https://doi.org/10.1007/978-3-030-31620-4_2

120. Montag, C., Sindermann, C., & Baumeister, H. (2020). Digital phenotyping in psychological and medical sciences: A reflection about necessary prerequisites to reduce harm and increase benefits. *Current Opinion in Psychology, 36*, 19–24. https://doi.org/10.1016/j.copsyc.2020.03.013
121. Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., Hirsch, D. A., Cooper, A. D., Drummond, S. P. A., & Markwald, R. R.: Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep, 44*(5), zsaa291. https://doi.org/10.1093/sleep/zsaa291
122. Driller, M., McQuillan, J., & O'Donnell, S. (2016). Inter-device reliability of an automatic-scoring actigraph for measuring sleep in healthy adults. *Sleep Science, 9*(3), 198–201. https://doi.org/10.1016/j.slsci.2016.08.003
123. Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems, 37*(46), 911–948. https://doi.org/10.17705/1CAIS.03746
124. Dimoka, A. (2012). How to conduct a functional magnetic resonance (fMRI) study in social science research. *MIS Quarterly, 36*(3), 811–840. https://doi.org/10.2307/41703482
125. Patron, E., Messerotti Benvenuti, S., Favretto, G., Gasparotto, R., & Palomba, D. (2014). Depression and reduced heart rate variability after cardiac surgery: The mediating role of emotion regulation. *Autonomic Neuroscience, 180*, 53–58. https://doi.org/10.1016/j.autneu.2013.11.004
126. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability in NeuroIS research: Review of empirical results. In: F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 285–299). Springer. https://doi.org/10.1007/978-3-031-13064-9_29
127. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A review of NeuroIS research with a focus on applied methods. In: F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 269–283). Springer. https://doi.org/10.1007/978-3-031-13064-9_28
128. Boucsein, W., & Thum, M. (1997). Design of work/rest schedules for computer work based on psychophysiological recovery measures. *International Journal of Industrial Ergonomics, 20*(1), 51–57. https://doi.org/10.1016/S0169-8141(96)00031-5
129. Riedl, R. (2009). Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen. *NeuroPsychoEconomics, 4*(1), 32–44.
130. Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G. R., Benbasat, I., Gefen, D., Kenning, P. H., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702. https://doi.org/10.2307/41703475

Neurophysiological Measurements in the Research Field of Interruption Science: Insights into Applied Methods for Different Interruption Types Based on an Umbrella Review

Fabian J. Stangl and René Riedl

Abstract Interruptions of various types, such as breaks, distractions, interventions, or intrusions, are ubiquitous in our daily lives. Interruption science is an interdisciplinary research field dedicated to the systematic investigation of interruptions, which have become a prevalent phenomenon in the economy and society in recent years. To lay the foundation for a better understanding of human neurophysiology related to the human perception of interruptions, we conducted an umbrella review to examine the applied neurophysiological measurements for different interruption types. We identified 72 empirical studies using a rigorous literature search process. Our analyses revealed three main measurement domains (brain imaging methods, autonomic nervous system activity measurements, and hormone measurements). We describe these three domains with respect to the applied neurophysiological measurements and four interruption types (break, distraction, intervention, and intrusion). Overall, this review provides methodological insights to advance interruption science research from a neurophysiological perspective.

Keywords Interruption Science · Interruption Science Research · Applied Neurophysiological Measurements · NeuroIS · Umbrella Review

F. J. Stangl (✉) · R. Riedl
Digital Business Institute, School of Business and Management, University of Applied Sciences Upper Austria, Steyr, Austria
e-mail: fabian.stangl@fh-steyr.at

R. Riedl
e-mail: rene.riedl@fh-steyr.at

R. Riedl
Institute of Business Informatics—Information Engineering, Johannes Kepler University Linz, Linz, Austria

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_11

1 Introduction

Interruption science is an interdisciplinary research field that systematically investigates the prevalent phenomenon of interruptions [1]. Interruptions can be defined as "*an unexpected suspension of the behavioral performance of, and/or attentional focus from, an ongoing work task*" [2, p. 817]. Researchers use interruption science to identify and mitigate predictors that affect the disruptiveness of interruptions, such as their length [3, 4]. By understanding the nature and impact of interruptions, guidelines and training programs can be developed to help individuals and organizations manage interruptions more effectively. This, in turn, improves primary task performance [5], overall work performance, productivity, and well-being [6, 7].

In recent years, the development and increasing use of technology in conjunction with global acceleration [8] has led to an increased focus on interruptions caused by digital technologies, also referred to as IT-mediated interruptions, in both scientific research and practice. Digital technologies have not only increased frequency [9, 10] but also the variety of interruptions that individuals encounter in their daily lives. The various interruption types include, for example, different modalities (e.g., auditory, pictographic, textual, visual; [11]), notification types (e.g., alerts, alarms, warnings, calls, summons; [12]), and transmission types (e.g., e-mail, instant messaging, telephone [13]). Specific interruption types may have negative effects on task performance and psychological well-being. For example, Chen and Karahanna [14] examined the effects of interruptions caused by different types of technologies. They found that work-related interruptions during leisure time caused by phone calls or messages can generate negative outcomes through interruption overload, whereas e-mail can lead to both positive and negative outcomes through task closure and psychological transition (between work and nonwork). Another study found that even brief interruptions (i.e., system response delays of 0.5–2.0 s) can have a negative effect on task performance [15, 16], and that the effects of interruptions can be cumulative, leading to longer-term consequences such as decreased productivity [17], increased number of errors [18], and fatigue [19]. Research on interruptions is therefore critical to better understand task performance, to work effectively and efficiently [20, 21], and to identify negative effects on psychological well-being [22].

A recent review of the conceptual and descriptive structure of interruptions revealed various attributes of interruptions [1]. To extend and complement this interdisciplinary overview of interruptions, the current paper aims to provide a concise overview of applied neurophysiological measurements for investigating human perceptions of interruptions. Neurophysiological measurements can provide a better understanding of human cognitive and affective processes and, for example, reveal how and why they experience certain effects when using digital technologies [23, 24]. As such, it has great potential for improving our understanding of the nature and impact of interruptions. Thus, the aim of this review is to lay the foundation for a better understanding of human neurophysiology during interruptions by using neurophysiological measures to examine human perception of interruptions. In addition, such an overview is likely to contribute to a better understanding of interruptions and, consequently, have a positive impact on interruption science itself by providing insights into the appropriate use of various applied methods [25, 26]. To the best of our knowledge, a review of the applied neurophysiological measurements of interruption perception has not yet been conducted. Specifically, we address the following research question in this review: **What neurophysiological measurements have been applied to examine different interruption types?**

The remainder of this paper is structured as follows. Section 2 outlines neurophysiological measurements of human perception. The knowledge presented in this section summarizes the main methodological approaches used to provide a brief overview. Section 3 describes the methodology used in this review. The results are presented in Sect. 4. Section 5 presents a discussion of the contributions and implications, along with the limitations of our review. Finally, in Sect. 6, we provide concluding remarks.

2 Neurophysiological Measurements of Human Perception

Neurophysiology is a scientific field devoted to studying the anatomy and functioning of the nervous system, including the brain and other neural tissues in the body [27]. At a high level of abstraction, human neurophysiology can be differentiated between the Central Nervous System (CNS) and the Peripheral Nervous System (PNS) as part of the human nervous system. The CNS can be further divided into neural tissues within the brain and spinal cord, while the PNS encompasses all other neural tissues outside of the CNS. The PNS can be further subdivided into the Somatic Nervous System (SNS) and Autonomic Nervous System (ANS). The SNS consists of cranial and spinal nerves that connect to sensory organs, muscles, joints, and skin, and its primary functions are movement production and sensory information transmission, such as touch or temperature [27, 28]. The ANS comprises a collection of nerves and nerve cells that play a crucial role in the innervation of various organs, including the respiratory tract, blood vessels, heart, urogenital organs, and intestines and can be divided into three components: the sympathetic division (which activates the body), the parasympathetic division (which relaxes the body) and the enteric nervous system (which governs the function of the gastrointestinal tract) [27–29]. In addition to CNS and PNS measurements, hormones (e.g., from blood, saliva, or urine samples) also have significant behavioral relevance (e.g., in stressful or trusting situations) [6, 28]. For example, research has shown that the perception of computer system breakdown leads to a significant feeling of stress, which has been measured by the stress hormone cortisol through saliva [30]. From an Information Systems (IS) perspective, the brain (i.e., the information processing unit), the sympathetic and parasympathetic divisions of the ANS (which maintain homeostasis in the body), and hormones (for behavioral relevance) are of critical importance [6, 28]. Therefore, neurophysiological measurements can examine human perception at different levels, ranging from individuals to groups, organizations, and societies [31]. Thus, they are of utmost importance in the research field of interruption science.

3 Review Methodology

To synthesize accumulated research on applied neurophysiological measurements in interruption science, we conducted an umbrella review [32–34] according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [35, 36]. Considering that the literature on interruption is scattered and poorly integrated across different scientific disciplines and research fields [1, 2] and that various terms exist with which interruptions are described (e.g., break, disruption, distraction, interruption, intrusion, suspension; [1]), an overarching review paper that consolidates the findings of previous literature reviews is an appropriate research method to obtain a comprehensive overview of applied neurophysiological measurements [37]. Theoretically, however, our literature review targeted two major streams of literature. First, we focused on the literature on digital stress, also referred to as technostress [38–41], where interruptions are one of the major triggers of digital stress that people must cope with in an increasingly digital world [1, 2, 14, 42, 43]. Indeed, empirical research found that office workers perceive about 70 interruptions per day from digital devices and programs used while performing actual work tasks [10], exemplifying that the proliferation of digital technologies in the workplace contributed dramatically to work-related interruptions [20]. Second, we also reviewed the literature on Neuro-Information Systems (NeuroIS), an interdisciplinary research field at the nexus of neurophysiology and digital technologies within the IS discipline that uses neuroscience and neurophysiological tools and methods to better understand human cognition, emotion, and behavior in IS contexts [23, 24, 28, 44–46]. As described below, our review methodology involved four steps that resulted in a literature base of 72 empirical studies that applied neurophysiological measurements to examine human perceptions of interruptions. Figure 1 graphically summarizes the flowchart of literature search and paper selection.

Step 1: Literature Identification—The starting point of our literature search was a publication on the biology of technostress. In this landmark publication, Riedl [6] reviewed and analyzed existing technostress research based on biological approaches from several different disciplines, and proposed a research agenda for future technostress research that also incorporates biology as an objective measurement method for determining human behavior toward digital technologies. We then conducted a forward search (i.e., citation tracking) based on this initial article using Google Scholar, which yielded 289 hits. For our literature search, we also considered reviews published in major academic databases in combination with physiological keywords that offer an introduction to the field of NeuroIS [23, 24, 28, 44] to identify further reviews on neurophysiological measurement methods to examine human behavior toward digital technologies (i.e., Google Scholar, Scopus, and Web of Science). However, we did not find any additional literature reviews, so we proceeded with our literature base, which consisted of 28 identified literature reviews (i.e., [6, 16, 44, 47–71]).

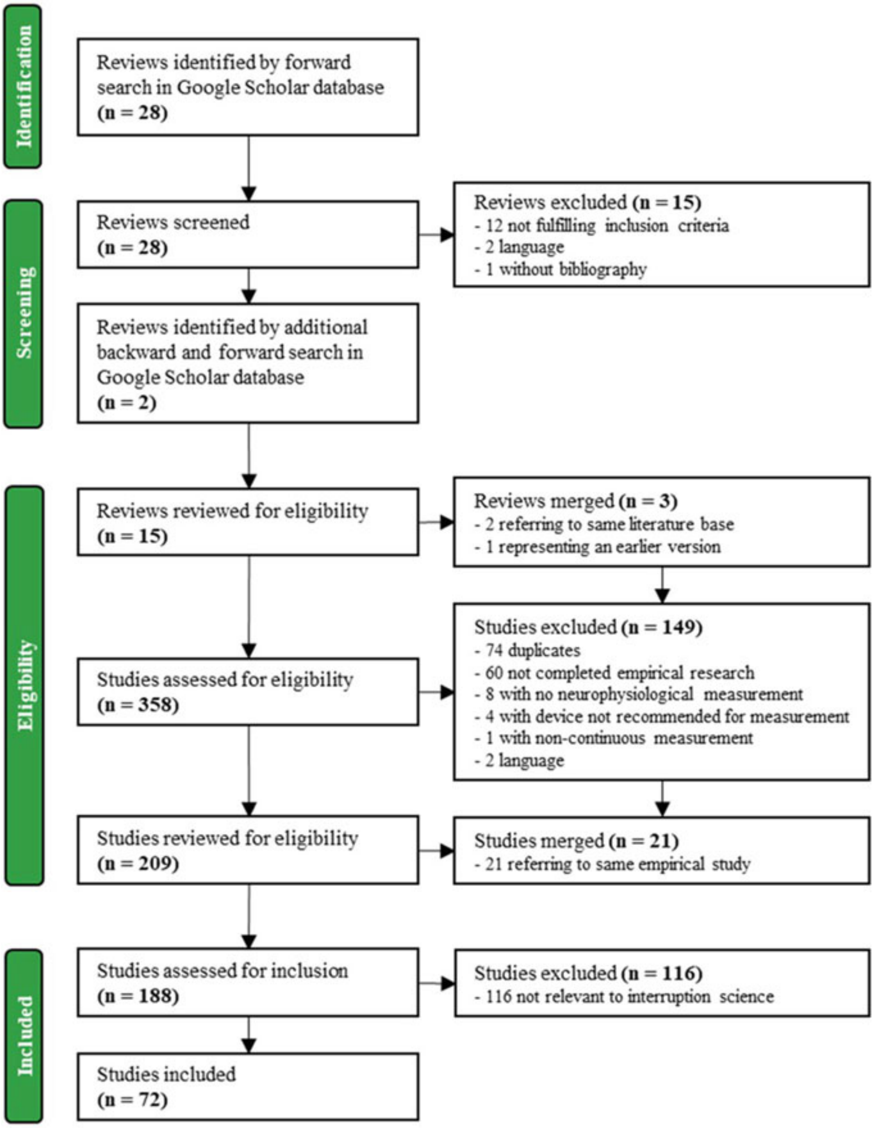


Fig. 1 PRISMA flowchart of literature search and paper selection

Step 2: Literature Screening—To be included in this review, we focused exclusively on peer-reviewed reviews of studies that empirically investigated human behavior toward digital technologies using neurophysiological measurements. Therefore, we excluded unrelated reviews based on the full text, leaving 13 reviews. For example, we removed 12 reviews that did not fulfill our inclusion criteria because they did not examine or only superficially examined applied neurophysiological measurement methods (9 in total; i.e., [52–54, 58, 60, 67–69, 71]) or that examined applied neurophysiological measurement methods in a research context not relevant to our research goal (3 in total; i.e., [50, 51, 61]). We also excluded reviews that were not written in English (2 in total; i.e., [48, 49]) or reviews that did not provide a bibliography of reviewed papers (1 in total; i.e., [55]). Following established recommendations for conducting literature searches [72, 73], the identified 13 reviews were then used for a backward search (i.e., searching the references) and again for a forward search using Google Scholar until we found no more relevant reviews. This procedure allowed us to identify two additional literature reviews (i.e., [74, 75]), resulting in a total of 15 reviews, including six relevant technostress reviews (i.e., [6, 16, 65, 66, 70, 74]) and nine relevant reviews with a neurophysiological measurement context related to human behavior toward digital technologies (i.e., [44, 47, 56, 57, 59, 62–64, 75]).

Step 3: Literature Eligibility—Before analyzing the collected papers, we merged reviews that used the same literature base to answer different research questions. In particular, we merged Stangl and Riedl [62–64], who reviewed the literature on heart rate and heart rate variability in NeuroIS to analyze the applied methods, results of empirical research, and reliability and validity of wearable devices (e.g., smartwatches) used in empirical research. We also merged Riedl et al. [44, 47], who initially presented an earlier version of their journal paper (i.e., [44]) at a conference (i.e., [47]). We then removed duplicates from our entire literature database (i.e., 12 reviews), leaving 284 unique papers out of 358 papers originally found. These papers were then analyzed in-depth based on the full text to ensure that the keywords were not superficial and that the article contained content relevant to our research goal (i.e., applied neurophysiological measurements to examine human perception of interruptions). Consequently, we excluded papers that had not yet completed their empirical research (60 papers in total; e.g., [76]), papers that did not apply neurophysiological measurements in their empirical research (8 papers in total; e.g., [77]), papers that conducted an empirical research with a device that, based on current knowledge and validation studies, are not recommended for physiological measurements for consumer, clinical, or research purposes (4 papers in total; e.g., [78]), papers where the physiological measurement could not be continuously measured in their empirical research (1 paper in total; [79]), and papers that were not written in English (2 papers in total; e.g., [80]), resulting in 209 unique papers remaining for further analysis.

In the next step, we merged the studies referring to the same empirical study. Here, we merged 39 papers [81–119], of which we considered 18 papers to be a more comprehensive version for further analysis [88–92, 94–102, 105, 107, 111, 112], resulting in 188 unique papers. Comparing the quantity of papers included in previous literature reviews, however, our umbrella-based review methodology also ensures a broad and valid consideration of applied neurophysiological measurement methods to examine human behavior toward digital technologies. Indeed, recent technostress reviews integrated and synthesized empirical data from 172 studies [58] or 102 studies [53], and another review on the development of NeuroIS research found 111 empirical studies that applied neurophysiological measurements [44].

Step 4: Literature Inclusion—To analyze the literature regarding interruptions as a trigger of digital stress [1, 2, 14, 42, 43], we drew on the interruption types outlined by Stangl and Riedl [43], which are described below. These interruption types are based on the typology of work interruptions proposed by Jett and George [120] and a taxonomy of IT-mediated interruptions proposed by Addas and Pinsonneault [10].

- **Break**: Endogenous event that results in time off from work to accommodate personal needs and the rhythm of the day (e.g., taking time to have lunch).
- **Distraction**: Exogenous event that triggers psychological reactions due to external stimuli or secondary activities that interrupt focused attention on primary task activities (e.g., individual is exposed to noise while performing a task).
- **Intervention**: Exogenous event that reveals a discrepancy between task performance expectations and the actual performance of primary task activities (e.g., an individual receives feedback relevant to task performance).
- **Intrusion**: Exogenous event that temporarily interrupts the execution of primary task activities due to an event that is irrelevant to task performance (e.g., unscheduled phone call).

This theory-oriented and rule-bounded procedure ensured that only relevant literature was included in the results, thereby excluding 116 papers that were not relevant to interruption science (i.e., [88–92, 94–97, 99–102, 105, 111, 121–221]). In the remaining 72 unique papers, though, neurophysiological measurements were applied to examine different interruption types (i.e., [30, 46, 98, 107, 112, 222–288]). Consequently, the literature base of our literature review comprises 72 empirical studies, including 51 journal papers (71%), 20 conference proceedings papers (28%), and 1 magazine article (1%).

4 Review Results

This section presents the main findings of the literature review. Our literature search revealed 72 empirical studies that applied neurophysiological measurements to examine different interruption types. As described in the following subsections, our analyses revealed three main research domains (brain imaging methods, ANS activity measurements, and hormone measurements), which are described with respect to applied neurophysiological measurements and interruption types. Note that this classification is based on an overview of the applied neurophysiological tools in NeuroIS research by Riedl and Léger [28, pp. 47–72]. Overall, our review shows that most empirical studies used measurements related to ANS activity (73%), whereas a smaller proportion relied on brain imaging methods (22%) or hormone measurements (5%).

Table 1 Examination of interruption perception with brain imaging methods

Measurement	Studies	Break	Distraction	Intervention	Intrusion
fMRI	11	N/A	2	5	4
EEG	9	N/A	3	2	5

Examination of Interruption Perception with Brain Imaging Methods

Seventeen papers were identified. These 17 papers included 11 studies using a functional magnetic resonance imaging scanner (fMRI; [46, 226, 233, 250, 253, 255, 273, 279, 287]) and 9 studies using electroencephalograms (EEG; [98, 241, 242, 247, 263, 266, 283, 285]). Note that the number of studies also considered papers in which multiple studies were conducted. For example, Züger and Fritz [285] conducted two EEG studies.

The analysis of the interruption types revealed that most interruptions could be classified as intrusion (43%; i.e., [98, 241, 250, 266, 273, 279, 285, 287]), followed by intervention (33%; i.e., [233, 253, 255, 263, 266]), and distraction (24%; i.e., [46, 226, 242, 247, 283]). As an example, Kalgotra et al. [98] gave their subjects reading tasks and exposed them to a series of randomly timed audio interruptions, which we classified as an intrusion. Note that the counting of interruption types also considered the papers in which multiple studies were conducted. For example, Kohrs, et al. [255] conducted 3 fMRI studies with to examine interventions (i.e., delayed feedback during task performance), which we counted as 3 interventions. Table 1 provides an overview of the interruption types measured using brain imaging methods.

Examination of Interruption Perception with ANS Activity Measurements

Fifty-six papers were identified. These 56 papers included 33 studies with eye-related measurements [223–225, 228–232, 236–239, 245, 248, 249, 259–262, 265, 266, 269, 270, 279, 281, 282, 284, 285, 288], 28 studies with heart-related measurements [112, 222, 227, 234, 235, 241, 243, 244, 246, 251, 252, 254, 256–258, 266, 268, 272, 274–278, 280, 285, 286], 18 studies with skin- or body-related measurements [107, 112, 241, 243, 252, 254, 257, 258, 264, 266, 271, 272, 275, 276, 278, 280, 285], 3 studies with muscle-related measurements [234, 235, 276], and 1 study with respiratory-related measurements [276]. Note that the number of studies also considered papers in which multiple studies were conducted. For example, Bahr and Ford [223] conducted two studies using eye-related measurements.

The analysis of the interruption types revealed that most interruptions could be classified as an intrusion (50%; i.e., [30, 223, 234, 235, 240, 241, 243, 251, 252,

Table 2 Examination of interruption perception with ANS activity measurements

Measurement	Studies	Break	Distraction	Intervention	Intrusion
Eye-related	33	N/A	19	7	8
Heart-related	28	1	3	7	18
Skin- or body-related	18	N/A	2	4	13
Muscle-related	3	N/A	N/A	N/A	3
Respiratory-related	1	N/A	N/A	N/A	1

257, 258, 266, 268, 271, 272, 274–276, 278, 279, 284–286, 288]), followed by a distraction (28%; i.e., [107, 224, 225, 227–232, 237, 238, 245, 246, 248, 249, 259–261, 267, 270, 280, 282]), an intervention (21%; i.e. [112, 222, 236, 239, 244, 248, 254, 264–266, 269, 277]), and a break (1%; i.e. [256]). As an example, Kuhmann [257] varied the system response time during task performance, which we classified as an intrusion. Note that the counting of interruption types also considered the papers in which multiple studies were conducted. For example, Jay et al. [249] conducted two studies with eye-related measurements to examine distraction (i.e., the effects of dynamic micro-content on attention allocation during browsing), which we counted as two distractions. Table 2 provides an overview of the interruption types measured using ANS activity measurements.

Examination of Interruption Perception with Hormone Measurements

Four papers were identified. These 4 papers included 4 studies using hormones from saliva samples [30, 240, 267] and 1 study using hormones from urine samples [251]. Note that the number of studies also considered papers in which multiple studies were conducted. For example, Galluch et al. [240] conducted two studies with hormone measurements.

The analysis of the interruption types revealed that most interruptions could be classified as an intrusion (80%; i.e., [30, 240, 251]), followed by a distraction (20%; i.e., [267]). As an example, Nomura et al. [267] whether an auditory stimulus has a post-task effect, which we classified as distraction. Note that the counting of interruption types also considered the papers in which multiple studies were conducted. For example, Galluch et al. [240] conducted 2 studies using hormones from saliva samples to examine intrusions (i.e., IT-mediated interruptions during task performance), which we counted as 2 intrusions. Table 3 provides an overview of the interruption types measured using hormone measurements.

Table 3 Examination of interruption perception with hormone measurements

Measurement	Studies	Break	Distraction	Intervention	Intrusion
Saliva samples	4	N/A	1	N/A	3
Urine samples	1	N/A	N/A	N/A	1

5 Review Discussion

Implications and Contributions

We contribute to research by providing a perspective on the applied methods to examine human perceptions of interruptions. Overall, our review indicates that eye-related measurements were the most applied neurophysiological measurement method in empirical studies (31%) for examining human perception of interruptions, followed by heart-related measurements (26%), and skin- or body-related measurements (17%). Only a small proportion of studies used fMRI (10%) and EEG (8%) to measure brain imaging and other measurements to measure ANS activity, such as muscle-related measurements (3%), respiratory-related measurements (1%), or hormone measurements (5%). Figure 2 graphically summarizes the main findings of our literature review on applied neurophysiological measurement methods to examine human perceptions of interruptions.

The main implication of this review is that the interdisciplinary research field of interruption science is a nurturing ground for the application of different neurophysiological measurement methods to the systematic study of interruptions rather than focusing on one measurement method. Methodologically speaking, NeuroIS research typically merges data collected from neuropsychological measurements with self-reported data to delve deeper into the cognitive and emotional processes of users during their interactions with digital technologies [23, 24, 28, 45, 289, 290] (for an overview of neurophysiological tools with a discussion of the strengths and weaknesses of each measurement method per research setting, please see [28, pp. 47–72]; for a detailed discussion of methods used in cognitive neuroscience, please see

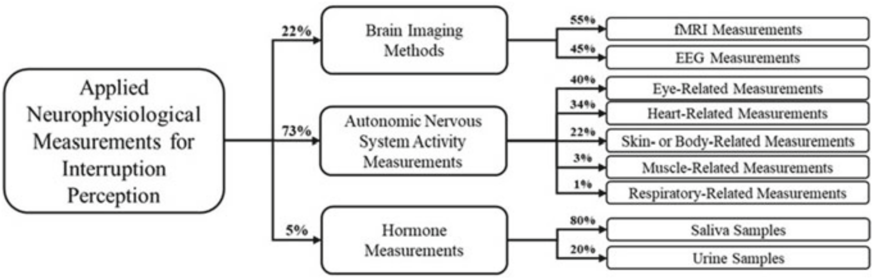


Fig. 2 Applied neurophysiological measurements for interruption perception

[291]). As an example, Galluch et al. [240] examined the impact of the quantity and content of IT-mediated interruptions (i.e., intrusion) with alpha-amylase, a salivary hormone that is an objective indicator of stress [292], along with self-reported data. Neurophysiological measurements can provide complementary insights to deepen understanding [24] and can encompass multiple measurement methods rather than being limited to a single approach. To assess the state of a knowledge worker's interruptibility with high accuracy, Züger and Fritz [285] combined a brain imaging method with multiple ANS activity measurements. Hence, conducting studies with neurophysiological measurements to examine human perceptions of interruptions seems promising for advancing interruption science research. Overall, focusing on a particular measurement method or a combination of different neurophysiological measurement methods offers promising opportunities for future research to explore and systematically examine human perceptions of interruptions.

Review Limitation

In our review methodology, we synthesized only empirical studies of reviews and did not conduct our own database searches to include empirical studies of applied neurophysiological measurements that examined human perceptions of interruptions. However, to the best of our knowledge, this is the most comprehensive review of human behavior toward digital technologies using neurophysiological measurements that currently exists. However, there are still empirical studies that have not been included in this review because of the review methodology (i.e., umbrella review following the PRIMSA checklist). For example, our review does not include the study by Chen et al. [293], who examined the neural mechanism of the brain during an interruption. Thus, future research addressing this limitation and extending our research findings could provide further insights into applied neurophysiological measurements to examine human perceptions of interruptions.

6 Concluding Remarks

Digital technologies shape everyday life. However, they have also given rise to a new type of problem: interruptions. In the workplace, they are noticeable, among other things, through long and variable response times [252, 258]. Moreover, the effects of digital interruptions are not limited to work-related tasks; they can also disrupt personal lives and social interactions. For example, a recent review found that (over-)use of Facebook as a social networking site poses both psychological and physiological risks [294]. These exemplary research findings underscore the particular value of exploring the use and role of digital technologies to counteract this "dark side" [295] and avoid constantly distracting or losing focus on the task at hand. As a foundation for a better understanding of human neurophysiology during interruptions, in this umbrella review, we addressed the question of what neurophysiological measurements can be used to examine different interruption types. Our analysis shows that different neurophysiological measurements are suitable for the research field of interruption science [1]. Therefore, we hope to see a growing number of further publications in this highly relevant research field in the coming years, as well as encouragement from researchers to advance scientific discourse on interruptions through further contributions to both scientific research and practice.

Acknowledgements This research was funded by the Austrian Science Fund (FWF) as part of the project "Technostress in Organizations" (project number: P 30865) and by the Federal State of Upper Austria through the Austrian Research Promotion Agency (FFG) as part of the project "Interruption" at the University of Applied Sciences Upper Austria.

References

1. Stangl, F. J., & Riedl, R. (2023). Interruption science as a research field: Towards a taxonomy of interruptions as a foundation for the field. *Frontiers in Psychology, 14*, 1043426. https://doi.org/10.3389/fpsyg.2023.1043426
2. Puranik, H., Koopman, J., & Vough, H. C. (2020). Pardon the interruption: An integrative review and future research agenda for research on work interruptions. *Journal of Management, 46*(6), 806–842. https://doi.org/10.1177/0149206319887428
3. Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26*(1), 39–83. https://doi.org/10.1207/s15516709cog2601_2
4. Trafton, J. G., & Monk, C. A. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics, 3*(1), 111–126. https://doi.org/10.1518/155723408X299852
5. Cades, D. M., Trafton, J. G., & Boehm-Davis, D. A. (2006). Mitigating disruptions: Can resuming an interrupted task be trained? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, no. 3, pp. 368–371). https://doi.org/10.1177/154193120605000333
6. Riedl, R. (2013). On the biology of technostress: Literature review and research agenda. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 44*(1), 18–55. https://doi.org/10.1145/2436239.2436242
7. Pflügner, K., Maier, C., Mattke, J., & Weitzel, T. (2021). Personality profiles that put users at risk of perceiving technostress: A qualitative comparative analysis with the big five personality traits. *Business & Information Systems Engineering, 63*(4), 389–402. https://doi.org/10.1007/s12599-020-00668-7
8. Couffe, C., & Michael, G. A. (2017). Failures due to interruptions or distractions: A review and a new framework. *The American Journal of Psychology, 130*(2), 163–181. https://doi.org/10.5406/amerjpsyc.130.2.0163
9. Mirhoseini, S., Hassanein, K., Head, M., & Watter, S. (2020). User performance in the face of IT interruptions: The role of executive functions. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019*. LNISO (Vol. 32, pp. 41–51). Springer. https://doi.org/10.1007/978-3-030-28144-1_5
10. Addas, S., & Pinsonneault, A. (2015). The many faces of information technology interruptions: A taxonomy and preliminary investigation of their performance effects. *Information Systems Journal, 25*(3), 231–273. https://doi.org/10.1111/isj.12064

11. Warnock, D., McGee-Lennon, M. R., & Brewster, S. (2011). The impact of unwanted multimodal notifications. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (pp. 177–184). https://doi.org/10.1145/2070481.2070510
12. Licoppe, C. (2010). The 'crisis of the summons': A transformation in the pragmatics of 'notifications', from phone rings to instant messaging. *The Information Society, 26*(4), 288–302.
13. Brixey, J. J., Walji, M., Zhang, J., Johnson, T. R., & Turley, J. P. (2004). Proposing a taxonomy and model of interruption. In *Proceedings of the 6th International Workshop on Enterprise Networking and Computing in Healthcare Industry* (pp. 184–188). https://doi.org/10.1109/HEALTH.2004.1324514
14. Chen, A. J., & Karahanna, E. (2018). Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS Quarterly, 42*(4), 1023–1042. https://doi.org/10.25300/MISQ/2018/13631
15. Monk, C. A., & Kidd, D. G. (2008). The effects of brief interruptions on task resumption. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 52, no. 4, pp. 403–407). https://doi.org/10.1177/154193120805200443
16. Riedl, R., & Fischer, T. (2018). System response time as a stressor in a digital world: Literature review and theoretical model. In F. F.-H. Nah & B. S. Xiao (Eds.), *HCI in Business, Government, and Organizations: 5th International Conference, HCIBGO 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings*. LNISA (Vol. 10923, pp. 175–186). Springer. https://doi.org/10.1007/978-3-319-91716-0_14
17. Baethge, A., Rigotti, T., & Roe, R. A. (2015). Just more of the same, or different? An integrative theoretical framework for the study of cumulative interruptions at work. *European Journal of Work and Organizational Psychology, 24*(2), 308–323. https://doi.org/10.1080/1359432X.2014.897943
18. Chen, Y., Fang, W., Bao, H., & Guo, B. (2022). The effect of task interruption on working memory performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/00187208221139017
19. Chen, Y., Fang, W., Guo, B., & Bao, H. (2022). The moderation effects of task attributes and mental fatigue on post-interruption task performance in a concurrent multitasking environment. *Applied Ergonomics, 102*, 103764. https://doi.org/10.1016/j.apergo.2022.103764
20. Tams, S., Ahuja, M., Thatcher, J. B., & Grover, V. (2020). Worker stress in the age of mobile technology: The combined effects of perceived interruption overload and worker control. *Journal of Strategic Information Systems, 29*(1), 101595. https://doi.org/10.1016/j.jsis.2020.101595
21. Baham, C., Kalgotra, P., Nasirpouri Shadbad, F., & Sharda, R. (2022). Generational differences in handling technology interruptions: A qualitative study. *European Journal of Information Systems,* 1–21. https://doi.org/10.1080/0960085X.2022.2070557
22. Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., Guo, Y., Harris, D., Ioannou, A., Jeong, K. (Kate), Konomi, S., Krömker, H., Kurosu, M., Lewis, J. R., Marcus, A., … Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction, 35*(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259
23. Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P. H., Müller-Putz, G. R., Pavlou, P. A., Straub, D. W., vom Brocke, J., & Weber, B. (2010). On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. *Communications of the Association for Information Systems, 27*(15), 243–264. https://doi.org/10.17705/1CAIS.02715
24. Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G. R., Benbasat, I., Gefen, D., Kenning, P. H., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702. https://doi.org/10.2307/41703475
25. Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems, 10*(2), 75–105. https://doi.org/10.1080/07421222.1993.11518001

26. Chong, S. W., & Reinders, H. (2021). A methodological review of qualitative research syntheses in CALL: The state-of-the-art. *System, 103*, 102646. https://doi.org/10.1016/j.system.2021.102646
27. Kandel, E. R., Koester, J. D., Mack, S. H., & Siegelbaum, S. A. (2021). *Principles of neural science*. McGraw Hill Professional.
28. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain*. Springer. https://doi.org/10.1007/978-3-662-45091-8
29. Gabella, G. (1976). *Structure of the autonomic nervous system*. Springer. https://doi.org/10.1007/978-94-009-5745-9
30. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2012). Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering, 4*(2), 61–69. https://doi.org/10.1007/s12599-012-0207-7
31. Riedl, R. (2009). Zum Erkenntnispotenzial der kognitiven Neurowissenschaften für die Wirtschaftsinformatik: Überlegungen anhand exemplarischer Anwendungen. *NeuroPsychoEconomics, 4*(1), 32–44.
32. Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008
33. Schryen, G., Benlian, A., Rowe, F., Gregor, S., Larsen, K., Petter, S., Paré, G., Wagner, G., Haag, S., & Yasasin, E. (2017). Literature reviews in IS research: What can be learnt from the past and other fields? *Communications of the Association for Information Systems, 41*(30), 759–774. https://doi.org/10.17705/1CAIS.04130
34. Schryen, G., Wagner, G., Benlian, A., & Paré, G. (2020). A knowledge development perspective on literature reviews: Validation of a new typology in the IS field. *Communications of the Association for Information Systems, 46*(7), 134–186. https://doi.org/10.17705/1CAIS.04607.
35. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
36. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007
37. Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare, 13*(3), 132–140. https://doi.org/10.1097/XEB.0000000000000055
38. Brod, C. (1984). *Technostress: The human cost of the computer revolution*. Addison-Wesley Publishing.
39. Weil, M. M., & Rosen, L. D. (1997). *TechnoStress: Coping with technology @work @home @play*. Wiley.
40. Ayyagari, R., Grover, V., & Purvis, R. (2011). Technostress: Technological antecedents and implications. *MIS Quarterly, 35*(4), 831–858. https://doi.org/10.2307/41409963
41. Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433. https://doi.org/10.1287/isre.1070.0165
42. Stangl, F. J., & Riedl, R. (2023). Digital Detoxing als Maßnahme gegen digitalen Stress: Auf die Verwendung digitaler Technologien temporär verzichten. *Theologisch-praktische Quartalschrift, 171*(3), 274–283.
43. Stangl, F. J., & Riedl, R. (2023). Interruptions in the workplace: An exploratory study among digital business professionals. In F. F.-H. Nah, & K. Siau (Eds.), *HCI in Business, Government and Organizations: 10th International Conference, HCIBGO 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part II*. LNCS (Vol. 14039, pp. 400–422). Springer. https://doi.org/10.1007/978-3-031-36049-7_29

44. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *ACM SIGMIS Database: The DATA BASE for Advances in Information Systems, 51*(3), 13–54. https://doi.org/10.1145/3410977.3410980
45. Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems, 15*(10), I–XXXV. https://doi.org/10.17705/1jais.00377
46. Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research, 22*(4), 687–702. https://doi.org/10.1287/isre.1100.0284
47. Riedl, R., Fischer, T., & Léger, P.-M. (2017). A decade of NeuroIS research: Status quo, challenges, and future directions. In *Proceedings of the 38th International Conference on Information Systems.*
48. Cuervo-Carabel, T., Orviz Martínez, N., Arce García, S., & Fernández Suárez, I. (2018). Tecnoestrés en la sociedad de la tecnología y la comunicación: Revisión bibliográfica a partir de la Web of Science. *Archivos de Prevención de Riesgos Laborales, 21*(1), 18–25. https://doi.org/10.12961/aprl.2018.21.01.4
49. Salazar-Concha, C., Ficapal-Cusí, P., & Boada-Grau, J. (2020). Tecnoestrés. Evolución del concepto y sus principales consecuencias. *Teuken Bidikay, 11*(17), 165–180. https://doi.org/10.33571/teuken.v11n17a9
50. Lux, E., Adam, M. T. P., Dorner, V., Helming, S., Knierim, M. T., & Weinhardt, C. (2018). Live biofeedback as a user interface design element: A review of the literature. *Communications of the Association for Information Systems, 43*(18), 257–296. https://doi.org/10.17705/1CAIS.04318
51. Kaltenegger, H. C., Becker, L., Rohleder, N., Nowak, D., & Weigl, M. (2021). Associations of working conditions and chronic low-grade inflammation among employees: A systematic review and meta-analysis. *Scandinavian Journal of Work, Environment & Health, 47*(8), 565–581. https://doi.org/10.5271/sjweh.3982
52. Kalischko, T., & Riedl, R. (2021). Electronic performance monitoring in the digital workplace: Conceptualization, review of effects and moderators, and future research opportunities. *Frontiers in Psychology, 12*, 633031. https://doi.org/10.3389/fpsyg.2021.633031
53. Nastjuk, I., Trang, S., Grummeck-Braamt, J.-V., Adam, M. T. P., & Tarafdar, M. (2023). Integrating and synthesising technostress research: A meta-analysis on technostress creators, outcomes, and IS usage contexts. *European Journal of Information Systems,* 1–22. https://doi.org/10.1080/0960085X.2022.2154712
54. Pflügner, K. (2022). Technostress management at the workplace: A systematic literature review. In *Proceedings of the 17th International Conference on Wirtschaftsinformatik*.
55. Grummeck-Braamt, J.-V., Nastjuk, I., Najmaei, A., & Adam, M. T. P. (2021). A bibliometric review of technostress: Historical roots, evolution and central publications of a growing research field. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 6621–6630).
56. Fischer, T., Halmerbauer, G., Meyr, E., & Riedl, R. (2018). Blood pressure measurement: A classic of stress measurement and its role in technostress research. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017*. LNISO (Vol. 25, pp. 25–35). Springer. https://doi.org/10.1007/978-3-319-67431-5_4
57. Fischer, T., & Riedl, R. (2015). The status quo of neurophysiology in organizational technostress research: A review of studies published from 1978 to 2015. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2015*. LNISO (Vol. 10, pp. 9–17). Springer. https://doi.org/10.1007/978-3-319-18702-0_2
58. Benzari, A., Khedhaouria, A., & Torrès, O. (2020). The rise of technostress: A literature review from 1984 until 2018. In *Proceedings of the 28th European Conference on Information Systems*.

59. Kalischko, T., & Riedl, R. (2020). Physiological measurement in the research field of electronic performance monitoring: Review and a call for NeuroIS studies. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2020*. LNISO (Vol. 43, pp. 233–243). Springer. https://doi.org/10.1007/978-3-030-60073-0_27
60. Salazar-Concha, C., Ficapal-Cusí, P., Boada-Grau, J., & Camacho, L. J. (2021). Analyzing the evolution of technostress: A science mapping approach. *Heliyon, 7*(4), e06726. https://doi.org/10.1016/j.heliyon.2021.e06726
61. Vogel, J., Auinger, A., & Riedl, R. (2019). Cardiovascular, neurophysiological, and biochemical stress indicators: A short review for information systems researchers. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2018*. LNISO (Vol. 29, pp. 259–273). Springer. https://doi.org/10.1007/978-3-030-01087-4_31
62. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A review of NeuroIS research with a focus on applied methods. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 269–283). Springer. https://doi.org/10.1007/978-3-031-13064-9_28
63. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability in NeuroIS research: Review of empirical results. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2022*. LNISO (Vol. 58, pp. 285–299). Springer. https://doi.org/10.1007/978-3-031-13064-9_29
64. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability with wearable devices: A systematic review. In *Proceedings of the 17th International Conference on Wirtschaftsinformatik*.
65. Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta—Techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal, 29*(1), 6–42. https://doi.org/10.1111/isj.12169
66. La Torre, G., Esposito, A., Sciarra, I., & Chiappetta, M. (2019). Definition, symptoms and risk of techno-stress: A systematic review. *International Archives of Occupational and Environmental Health, 92*(1), 13–35. https://doi.org/10.1007/s00420-018-1352-1
67. Marsh, E., Vallejos, E. P., & Spence, A. (2022). The digital workplace and its dark side: An integrative review. *Computers in Human Behavior, 128*, 107118. https://doi.org/10.1016/j.chb.2021.107118
68. Nisafani, A. S., Kiely, G., & Mahony, C. (2020). Workers' technostress: A review of its causes, strains, inhibitors, and impacts. *Journal of Decision Systems, 29*(sup1), 243–258. https://doi.org/10.1080/12460125.2020.1796286
69. Borle, P., Reichel, K., Niebuhr, F., & Voelter-Mahlknecht, S. (2021). How are techno-stressors associated with mental health and work outcomes? A systematic review of occupational exposure to information and communication technologies within the technostress model. *International Journal of Environmental Research and Public Health, 18*(16), 8673. https://doi.org/10.3390/ijerph18168673
70. Fischer, T., & Riedl, R. (2017). Technostress research: A nurturing ground for measurement pluralism? *Communications of the Association for Information Systems, 40*(17), 375–401. https://doi.org/10.17705/1CAIS.04017
71. Agogo, D., & Hess, T. J. (2018). "How does tech make you feel?" a review and examination of negative affective responses to technology use. *European Journal of Information Systems, 27*(5), 570–599. https://doi.org/10.1080/0960085X.2018.1435230
72. Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.
73. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In S. Newell, E. A. Whitley, N. Pouloudi, J. Wareham, & L. Mathiassen (Eds.), *Proceedings of the 17th European Conference on Information Systems* (pp. 2206–2217).

74. Boucsein, W. (2009). Forty years of research on system response times—What did we learn from it? In C. M. Schlick (Ed.), *Industrial Engineering and Ergonomics: Visions, Concepts, Methods and Tools Festschrift in Honor of Professor Holger Luczak* (pp. 575–593). Springer. https://doi.org/10.1007/978-3-642-01293-8_42
75. Weber, B., Fischer, T., & Riedl, R. (2021). Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review. *Journal of Systems and Software, 178*, 110946. https://doi.org/10.1016/j.jss.2021.110946
76. Büttner, R. (2017). Asking both the user's brain and its owner using subjective and objective psychophysiological NeuroIS instruments. In *Proceedings of the 38th International Conference on Information Systems.*
77. Aiello, J. R., & Kolb, K. J. (1995). Electronic performance monitoring and social context: Impact on productivity and stress. *Journal of Applied Psychology, 80*(3), 339–353.
78. Gaskin, J., Jenkins, J. L., Meservy, T. O., Steffen, J., & Payne, K. (2017). Using wearable devices for non-invasive, inexpensive physiological data collection. In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 597–605). https://doi.org/10.24251/HICSS.2017.072
79. Fischer, T., & Riedl, R. (2020). Technostress measurement in the field: A case report. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2020*. LNISO (Vol. 43, pp. 71–78). Springer. https://doi.org/10.1007/978-3-030-60073-0_9
80. Kuhmann, W., Schäfer, F., & Boucsein, W. (1990). Effekte von Wartezeiten innerhalb einfacher Aufgaben: Eine Analogie zu Wartezeiten in der Mensch-Computer-Interaktion. *Zeitschrift für experimentelle und angewandte Psychologie, 37*(2), 242–265.
81. Adam, M. T. P., Gamer, M., Krämer, J., & Weinhardt, C. (2011). Measuring emotions in electronic markets. In *Proceedings of the 32nd International Conference on Information Systems.*
82. Arnetz, B. B., & Wiholm, C. (1997). Technological stress: Psychophysiological symptoms in modern offices. *Journal of Psychosomatic Research, 43*(1), 35–42. https://doi.org/10.1016/S0022-3999(97)00083-4
83. Kalgotra, P., Sharda, R., & McHaney, R. (2016). Understanding the impact of interruptions on knowledge work: An exploratory neuroimaging study. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 658–667). https://doi.org/10.1109/HICSS.2016.87
84. Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2011). Trusting humans and avatars: Behavioral and neural evidence. In *Proceedings of the 32nd International Conference on Information Systems.*
85. Rodger, J. A., & Gonzalez, S. P. (2013). Emotion and memory in technology adoption and diffusion. In *Proceedings of the 19th Americas Conference on Information Systems.*
86. Wells, T. M., & Dennis, A. R. (2014). You said what? Assessing the impact of collaboration technologies and message characteristics using physiological measures. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 442–451). https://doi.org/10.1109/HICSS.2014.62
87. Yang, S.-F., & Lin, H.-H. (2014). Effects of attribute framing varying with the elaboration in online shopping: An eye-tracking approach. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3083–3092). IEEE. https://doi.org/10.1109/HICSS.2014.384
88. Adam, M. T. P., Krämer, J., & Weinhardt, C. (2012). Excitement up! Price down! Measuring emotions in Dutch auctions. *International Journal of Electronic Commerce, 17*(2), 7–40. https://doi.org/10.2753/JEC1086-4415170201
89. Arnetz, B. B., & Berg, M. (1996). Melatonin and adrenocorticotropic hormone levels in video display unit workers during work and leisure. *Journal of Occupational and Environmental Medicine, 38*(11), 1108–1110.
90. Berg, M., Arnetz, B. B., Lidén, S., Eneroth, P., & Kallner, A. (1992). Techno-stress. A psychophysiological study of employees with VDU-associated skin complaints. *Journal of Occupational Medicine, 34*(7), 698–701.

91. Couceiro, R., Castelhano, J., Castelo-Branco, M., Duarte, G., Duarte, C., de Carvalho, P., Durães, J., Teixeira, C., & Madeira, H. (2019). Pupillography as indicator of programmers' mental effort and cognitive overload. In *Proceedings of the 49th International Conference on Dependable Systems and Networks* (pp. 638–644). https://doi.org/10.1109/DSN.2019.00069
92. Crk, I., Kluthe, T., & Stefik, A. (2016). Understanding programming expertise: An empirical study of phasic brain wave changes. *ACM Transactions on Computer-Human Interaction, 23*(1), 1–29. https://doi.org/10.1145/2829945
93. Couceiro, R., Barbosa, R., Durães, J., Duarte, G., Castelhano, J., Duarte, C., Teixeira, C., Laranjeiro, N., Medeiros, J., de Carvalho, P., Castelo-Branco, M., & Madeira, H. (2019). Spotting problematic code lines using nonintrusive programmers' biofeedback. In *Proceedings of the 30th International Symposium on Software Reliability Engineering* (pp. 93–103). https://doi.org/10.1109/ISSRE.2019.00019
94. Doukakis, S., Papalaskari, M.-A., Vlamos, P., Plerou, A., & Giannopoulou, P. (2020). Undergraduate students' brain activity in visual and textual programming. In P. Vlamos (Ed.), *GeNeDis 2018: Computational Biology and Bioinformatics*. AEMB (Vol. 1194, pp. 425–435). Springer. https://doi.org/10.1007/978-3-030-32622-7_40
95. Fakhoury, S., Roy, D., Ma, Y., Arnaoudova, V., & Adesope, O. (2020). Measuring the impact of lexical and structural inconsistencies on developers' cognitive load during bug localization. *Empirical Software Engineering, 25*(3), 2140–2178. https://doi.org/10.1007/s10664-019-09751-4
96. Hu, Q., West, R. L., & Smarandescu, L. (2015). The role of self-control in information security violations: Insights from a cognitive neuroscience perspective. *Journal of Management Information Systems, 31*(4), 6–48. https://doi.org/10.1080/07421222.2014.1001255
97. Ishida, T., & Uwano, H. (2019). Synchronized analysis of eye movement and EEG during program comprehension. In *Proceedings of the 6th International Workshop on Eye Movements in Programming* (pp. 26–32). https://doi.org/10.1109/EMIP.2019.00012
98. Kalgotra, P., Sharda, R., & McHaney, R. (2019). Don't disturb me! Understanding the impact of interruptions on knowledge work: An exploratory neuroimaging study. *Information Systems Frontiers, 21*(5), 1019–1030. https://doi.org/10.1007/s10796-017-9812-9
99. Riedl, R., Mohr, P. N. C., Kenning, P. H., Davis, F. D., & Heekeren, H. R. (2014). Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems, 30*(4), 83–114. https://doi.org/10.2753/MIS0742-1222300404
100. Rodger, J. A. (2014). Reinforcing inspiration for technology acceptance: Improving memory and software training results through neuro-physiological performance. *Computers in Human Behavior, 38*, 174–184. https://doi.org/10.1016/j.chb.2014.05.049
101. Wells, T. M., & Dennis, A. R. (2016). To email or not to email: The impact of media on psychophysiological responses and emotional content in utilitarian and romantic communication. *Computers in Human Behavior, 54*, 1–9. https://doi.org/10.1016/j.chb.2015.07.036
102. Yang, S.-F. (2015). An eye-tracking study of the Elaboration Likelihood Model in online shopping. *Electronic Commerce Research and Applications, 14*(4), 233–240. https://doi.org/10.1016/j.elerap.2014.11.007
103. Lutz, B., Adam, M. T. P., Feuerriegel, S., Pröllochs, N., & Neumann, D. (2020). Affective information processing of fake news: Evidence from NeuroIS. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019*. LNISO (Vol. 32, pp. 121–128). Springer. https://doi.org/10.1007/978-3-030-28144-1_13
104. Couceiro, R., Castelhano, J., Castelo-Branco, M., Duarte, G., Duarte, C., de Carvalho, P., Durães, J., Teixeira, C., & Madeira, H. (2019). Biofeedback augmented software engineering: Monitoring of programmers' mental effort. In *Proceedings of the 41st International Conference on Software Engineering* (pp. 37–40). https://doi.org/10.1109/ICSE-NIER.2019.00018
105. Lutz, B., Adam, M. T. P., Feuerriegel, S., Pröllochs, N., & Neumann, D. (2020). Identifying linguistic cues of fake news associated with cognitive and affective processing: Evidence from NeuroIS. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T.

Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2020*. LNISO (Vol. 43, pp. 16–23). Springer. https://doi.org/10.1007/978-3-030-60073-0_2
106. Yang, E., & Dorneich, M. C. (2015). The effect of time delay on emotion, arousal, and satisfaction in human-robot interaction. In *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting* (pp. 443–447).
107. Yang, E., & Dorneich, M. C. (2017). The emotional, cognitive, physiological, and performance effects of variable time delay in robotic teleoperation. *International Journal of Social Robotics, 9*(4), 491–508. https://doi.org/10.1007/s12369-017-0407-x
108. Wastell, D. G., & Cooper, C. L. (1996). Stress and technological innovation: A comparative study of design practices and implementation strategies. *European Journal of Work and Organizational Psychology, 5*(3), 377–397. https://doi.org/10.1080/13594329608414866
109. Wastell, D. G., & Newman, M. (1996). Information system design, stress and organisational change in the ambulance services: A tale of two cities. *Accounting, Management and Information Technologies, 6*(4), 283–300. https://doi.org/10.1016/S0959-8022(96)90017-X
110. Wastell, D. G., & Newman, M. (1996). Information systems development in the ambulance service: A tale of two cities. In *Proceedings of the 4th European Conference on Information Systems* (pp. 539–551).
111. Wastell, D. G., & Newman, M. (1996). Stress, control and computer system design: A psychophysiological field study. *Behaviour & Information Technology, 15*(3), 183–192. https://doi.org/10.1080/014492996120247
112. Galletta, D. F., & Grant, R. (1995). Silicon supervisors and stress: Merging new evidence from the field. *Accounting, Management and Information Technologies, 5*(3–4), 163–183. https://doi.org/10.1016/0959-8022(96)00003-3
113. Huston, T. L., Galletta, D. F., & Huston, J. L. (1993). The effects of computer monitoring on employee performance and stress: Results of two experimental studies. In *Proceedings of the 26th Hawaii International Conference on System Sciences* (Vol. 4, pp. 568–574). https://doi.org/10.1109/HICSS.1993.284234
114. Crk, I., & Kluthe, T. (2014). Toward using alpha and theta brain waves to quantify programmer expertise. In *Proceedings of the 36th International Conference on Engineering in Medicine and Biology Society* (pp. 5373–5376). https://doi.org/10.1109/EMBC.2014.6944840
115. Crk, I., & Kluthe, T. (2016). Assessing the contribution of the individual alpha frequency (IAF) in an EEG-based study of program comprehension. In *Proceedings of the 38th International Conference on Engineering in Medicine and Biology Society* (pp. 4601–4604). https://doi.org/10.1109/EMBC.2016.7591752
116. Doukakis, S. (2019). Exploring brain activity and transforming knowledge in visual and textual programming using neuroeducation approaches. *AIMS Neuroscience, 6*(3), 175–190. https://doi.org/10.3934/Neuroscience.2019.3.175
117. Fakhoury, S., Ma, Y., Arnaoudova, V., & Adesope, O. (2018). The effect of poor source code lexicon and readability on developers' cognitive load. In *Proceedings of the 26th International Conference on Program Comprehension* (pp. 286–296). https://doi.org/10.1145/3196321.3196347
118. Hu, Q., West, R. L., Smarandescu, L., & Yaple, Z. (2014). Why individuals commit information security violations: Neural correlates of decision processes and self-control. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3234–3243).
119. Ishida, T., & Uwano, H. (2019). Time series analysis of programmer's EEG for debug state classification. In *Proceedings of the Conference Companion of the 3rd International Conference on Art, Science, and Engineering of Programming* (pp. 1–7). https://doi.org/10.1145/3328433.3328459
120. Jett, Q. R., & George, J. M. (2003). Work interrupted: A closer look at the role of interruptions in organizational life. *Academy of Management Review, 28*(3), 494–507. https://doi.org/10.5465/amr.2003.10196791
121. Vrzakova, H., Begel, A., Mehtätalo, L., & Bednarik, R. (2020). Affect recognition in code review: An in-situ biometric study of reviewer's affect. *Journal of Systems and Software, 159*, 110434. https://doi.org/10.1016/j.jss.2019.110434

122. Walden, E. A., Browne, G. J., & Oboyle, M. (2015). Computational thinking: Changes to the human connectome associated with learning to program. In *Proceedings of the 36th International Conference on Information Systems.*
123. Wang, Q., Meng, L., Liu, M., Wang, Q., & Ma, Q. (2016). How do social-based cues influence consumers' online purchase decisions? An event-related potential study. *Electronic Commerce Research, 16*(1), 1–26. https://doi.org/10.1007/s10660-015-9209-0
124. Wang, C.-C., & Hsu, M.-C. (2014). An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction. *Information & Management, 51*(7), 912–923. https://doi.org/10.1016/j.im.2014.05.010
125. Wang, H.-C., & Doong, H.-S. (2017). An exploratory study on consumers' attention towards social media advertising: An electroencephalography approach. In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 823–832).
126. Warkentin, M., Walden, E. A., Johnston, A. C., & Straub, D. W. (2016). Neural correlates of protection motivation for secure IT behaviors: An fMRI examination. *Journal of the Association for Information Systems, 17*(3), 194–215. https://doi.org/10.17705/1jais.00424
127. Weber, B., Pinggera, J., Neurauter, M., Zugal, S., Martini, M., Furtner, M., Sachse, P., & Schnitzer, D. (2016). Fixation patterns during process model creation: Initial steps toward neuro-adaptive process modeling environments. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 600–609). https://doi.org/10.1109/HICSS.2016.81
128. Wiholm, C., & Arnetz, B. B. (1997). Musculoskeletal symptoms and headaches in VDU users—a psychophysiological study. *Work & Stress, 11*(3), 239–250. https://doi.org/10.1080/02678379708256838
129. Wrzesien, M., Rodríguez, A., Rey, B., Alcañiz, M., Baños, R. M., & Vara, M. D. (2015). How the physical similarity of avatars can influence the learning of emotion regulation strategies in teenagers. *Computers in Human Behavior, 43*, 101–111. https://doi.org/10.1016/j.chb.2014.09.024
130. Wulff-Jensen, A., Ruder, K., Triantafyllou, E., & Bruni, L. E. (2019). Gaze strategies can reveal the impact of source code features on the cognitive load of novice programmers. In H. Ayaz & L. Mazur (Eds.), *Advances in Neuroergonomics and Cognitive Engineering: Proceedings of the AHFE 2018 International Conference on Neuroergonomics and Cognitive Engineering, July 21–25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida USA*. AISC (Vol. 775, pp. 91–100). Springer. https://doi.org/10.1007/978-3-319-94866-9_9
131. Yamamoto, A., Uwano, H., & Ikutani, Y. (2016). Programmer's electroencephalogram who found implementation strategy. In *Proceedings of the 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science* (pp. 164–168). IEEE. https://doi.org/10.1109/ACIT-CSII-BCD.2016.041
132. Yeh, M. K.-C., Gopstein, D., Yan, Y., & Zhuang, Y. (2017). Detecting and comparing brain activity in short program comprehension using EEG. In *Proceedings of the Frontiers in Education Conference* (pp. 1–5). https://doi.org/10.1109/FIE.2017.8190486
133. Yen, M.-H., & Wu, Y.-T. (2017). The role of university students' informal reasoning ability and disposition in their engagement and outcomes of online reading regarding a controversial issue: An eye tracking study. *Computers in Human Behavior, 75*, 14–24. https://doi.org/10.1016/j.chb.2017.04.054
134. Yuan, L. (Ivy), Dennis, A. R., & Potter, R. F. (2016). Interacting like humans? Understanding the neurophysiological processes of anthropomorphism and consumer's willingness to pay in online auctions. In *Proceedings of the 37th International Conference on Information Systems.*
135. Zhou, J., Sun, J., Chen, F., Wang, Y., Taib, R., Khawaji, A., & Li, Z. (2015). Measurable decision making with GSR and pupillary analysis for intelligent user interface. *ACM Transactions on Computer-Human Interaction, 21*(6), 1–23. https://doi.org/10.1145/2687924
136. Kim, N. W., Bylinskii, Z., Borkin, M. A., Gajos, K. Z., Oliva, A., Durand, F., & Pfister, H. (2017). BubbleView: An interface for crowdsourcing image importance maps and tracking

visual attention. *ACM Transactions on Computer-Human Interaction, 24*(5), 1–40. https://doi.org/10.1145/3131275

137. Ahonen, L., Cowley, B. U., Hellas, A., & Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Scientific Reports, 8*(1), 3138. https://doi.org/10.1038/s41598-018-21518-3
138. Ahonen, L., Cowley, B. U., Torniainen, J., Ukkonen, A., Vihavainen, A., & Puolamäki, K. (2016). Cognitive collaboration found in cardiac physiology: Study in classroom environment. *PLoS ONE, 11*(7), e0159178. https://doi.org/10.1371/journal.pone.0159178
139. Arnetz, B. B. (1996). Techno-stress: A prospective psychophysiological study of the impact of a controlled stress-reduction program in advanced telecommunication systems design work. *Journal of Occupational and Environmental Medicine, 38*(1), 53–65.
140. Aschwanden, C., & Crosby, M. E. (2006). Code scanning patterns in program comprehension. In *Proceedings of the 39th Hawaii International Conference on System Sciences*.
141. Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction, 14*(4), 1–28. https://doi.org/10.1145/1314683.1314689
142. Barral, O., Kosunen, I., & Jacucci, G. (2017). No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. *ACM Transactions on Computer-Human Interaction, 24*(6), 1–29. https://doi.org/10.1145/3157730
143. Behroozi, M., Lui, A., Moore, I., Ford, D., & Parnin, C. (2018). Dazed: Measuring the cognitive load of solving technical interview problems at the whiteboard. In *Proceedings of the 40th International Conference on Software Engineering* (pp. 93–96). https://doi.org/10.1145/3183399.3183415
144. Behroozi, M., & Parnin, C. (2018). Can we predict stressful technical interview settings through eye-tracking? In *Proceedings of the Workshop on Eye Movements in Programming* (pp. 1–5). https://doi.org/10.1145/3216723.3216729
145. Boucsein, W., & Thum, M. (1997). Design of work/rest schedules for computer work based on psychophysiological recovery measures. *International Journal of Industrial Ergonomics, 20*(1), 51–57. https://doi.org/10.1016/S0169-8141(96)00031-5
146. Büttner, R., Bachus, L., Konzmann, L., & Prohaska, S. (2019). Asking both the user's heart and its owner: Empirical evidence for substance dualism. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2018*. LNISO (Vol. 29, pp. 251–257). Springer. https://doi.org/10.1007/978-3-030-01087-4_30
147. Castelhano, J., Duarte, I. C., Ferreira, C., Duraes, J., Madeira, H., & Castelo-Branco, M. (2019). The role of the insula in intuitive expert bug detection in computer code: An fMRI study. *Brain Imaging and Behavior, 13*(3), 623–637. https://doi.org/10.1007/s11682-018-9885-1
148. Caya, O., Brunelle, É., Léger, P.-M., & Grebot, T. (2012). An empirical study on emotions, knowledge management processes, and performance within integrated business process teams. In *Proceedings of the 45th Hawaii International Conference on System Sciences* (pp. 514–524). https://doi.org/10.1109/HICSS.2012.98
149. Cipresso, P., Serino, S., Gaggioli, A., Albani, G., Mauro, A., & Riva, G. (2015). Psychometric modeling of the pervasive use of Facebook through psychophysiological measures: Stress or optimal experience? *Computers in Human Behavior, 49*, 576–587. https://doi.org/10.1016/j.chb.2015.03.068
150. Cole, M. J., Hendahewa, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. *ACM Transactions on Information Systems, 33*(1), 1–39. https://doi.org/10.1145/2699656
151. Comesaña, M., Soares, A. P., Perea, M., Piñeiro, A. P., Fraga, I., & Pinheiro, A. (2013). ERP correlates of masked affective priming with emoticons. *Computers in Human Behavior, 29*(3), 588–595. https://doi.org/10.1016/j.chb.2012.10.020
152. Dimoka, A. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly, 34*(2), 373–396. https://doi.org/10.2307/20721433

153. Djamasbi, S., Siegel, M., Skorinko, J., & Tullis, T. S. (2011). Online viewing and aesthetic preferences of generation Y and the baby boom generation: Testing user web site experience through eye tracking. *International Journal of Electronic Commerce, 15*(4), 121–158. https://doi.org/10.2753/JEC1086-4415150404
154. Djamasbi, S., Mehta, D., & Samani, A. (2012). Eye movements, perceptions, and performance. In *Proceedings of the 18th Americas Conference on Information Systems.*
155. Dogusoy-Taylan, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in Human Behavior, 36*, 82–93. https://doi.org/10.1016/j.chb.2014.03.036
156. Duraisingam, A., Palaniappan, R., & Andrews, S. (2017). Cognitive task difficulty analysis using EEG and data mining. In *Proceedings of the Conference on Emerging Devices and Smart Systems* (pp. 52–57). https://doi.org/10.1109/ICEDSS.2017.8073658
157. Eckhardt, A., Maier, C., Hsieh, J. J. P.-A., Chuk, T., Chan, A. B., Hsiao, J. H., & Büttner, R. (2013). Objective measures of IS usage behavior under conditions of experience and pressure using eye fixation data. In *Proceedings of the 34th International Conference on Information Systems.*
158. Etco, M., Sénécal, S., Léger, P.-M., & Fredette, M. (2017). The influence of online search behavior on consumers' decision-making heuristics. *Journal of Computer Information Systems, 57*(4), 344–352. https://doi.org/10.1080/08874417.2016.1232987
159. Fadel, K. J., Meservy, T. O., & Jensen, M. L. (2015). Exploring knowledge filtering processes in electronic networks of practice. *Journal of Management Information Systems, 31*(4), 158–181. https://doi.org/10.1080/07421222.2014.1001262
160. Floyd, B., Santander, T., & Weimer, W. (2017). Decoding the representation of code in the brain: An fMRI study of code review and expertise. In *Proceedings of the 39th International Conference on Software Engineering* (pp. 175–186). https://doi.org/10.1109/ICSE.2017.24
161. Fritz, T., Begel, A., Müller, S. C., Yigit-Elliott, S., & Züger, M. (2014). Using psychophysiological measures to assess task difficulty in software development. In *Proceedings of the 36th International Conference on Software Engineering* (pp. 402–413). https://doi.org/10.1145/2568225.2568266
162. Fucci, D., Girardi, D., Novielli, N., Quaranta, L., & Lanubile, F. (2019). A replication study on code comprehension and expertise using lightweight biometric sensors. In *Proceedings of the 27th International Conference on Program Comprehension* (pp. 311–322). https://doi.org/10.1109/ICPC.2019.00050
163. González, V. M., Robbes, R., Góngora, G., & Medina, S. (2015). Measuring concentration while programming with low-cost BCI devices: Differences between debugging and creativity tasks. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition: 9th International Conference, AC 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings*. LNCS (Vol. 9183, pp. 605–615). Springer. https://doi.org/10.1007/978-3-319-20816-9_58
164. González-Cabrera, J., Calvete Zumalde, E., León-Mejía, A. C., Pérez-Sancho, C., & Peinado, J. M. (2017). Relationship between cyberbullying roles, cortisol secretion and psychological stress. *Computers in Human Behavior, 70*, 153–160. https://doi.org/10.1016/j.chb.2016.12.054
165. Hariharan, A., Dorner, V., & Adam, M. T. P. (2017). Impact of cognitive workload and emotional arousal on performance in cooperative and competitive interactions. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2016*. LNISO (Vol. 16, pp. 35–42). Springer. https://doi.org/10.1007/978-3-319-41402-7_5
166. Hu, L., Zhang, W., & Xu, Q. (2013). The determinants of online payment method choice: Insight from an eyetracking study. In *Proceedings of the 12th Wuhan International Conference on e-Business* (pp. 434–443).
167. Huang, Y.-F., & Kuo, F.-Y. B. (2012). How impulsivity affects consumer decision-making in e-commerce. *Electronic Commerce Research and Applications, 11*(6), 582–590. https://doi.org/10.1016/j.elerap.2012.09.004

168. Huang, Y.-F., Kuo, F.-Y. B., & Lin, C. S. (2017). Behavior regulation in social media: A neuroscientific investigation. In *Proceedings of the 38th International Conference on Information Systems*.
169. Huang, Y.-F., Kuo, F.-Y. B., Luu, P., Tucker, D., & Hsieh, P.-J. (2015). Hedonic evaluation can be automatically performed: An electroencephalography study of website impression across two cultures. *Computers in Human Behavior, 49*, 138–146. https://doi.org/10.1016/j.chb.2015.02.016
170. Huang, Y., Liu, X., Krueger, R., Santander, T., Hu, X., Leach, K., & Weimer, W. (2019). Distilling neural representations of data structure manipulation using fMRI and fNIRS. In *Proceedings of the 41st International Conference on Software Engineering* (pp. 396–407). https://doi.org/10.1109/ICSE.2019.00053
171. Ikutani, Y., & Uwano, H. (2014). Brain activity measurement during program comprehension with NIRS. In *Proceedings of the 15th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (pp. 1–6). https://doi.org/10.1109/SNPD.2014.6888727
172. Ikutani, Y., Kubo, T., Nishida, S., Hata, H., Matsumoto, K., Ikeda, K., & Nishimoto, S. (2021). Expert programmers have fine-tuned cortical representations of source code. *eNeuro, 8*(1), ENEURO.0405-20.2020. https://doi.org/10.1523/ENEURO.0405-20.2020
173. Inuzuka, M., Tanaka, Y., & Tsubakimoto, M. (2017). Students' comprehension of scientific discussion: Using eye-tracking technique to investigate the effects of social-media messages on television. In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 2106–2115).
174. Ivanova, A. A., Srikant, S., Sueoka, Y., Kean, H. H., Dhamala, R., O'Reilly, U.-M., Bers, M. U., & Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general executive brain regions. *eLife, 9,* e58906. https://doi.org/10.7554/eLife.58906
175. Javor, A., Ransmayr, G., Struhal, W., & Riedl, R. (2016). Parkinson patients' initial trust in avatars: Theory and evidence. *PLoS ONE, 11*(11), e0165998. https://doi.org/10.1371/journal.pone.0165998
176. Jbara, A., & Feitelson, D. G. (2017). How programmers read regular code: A controlled experiment using eye tracking. *Empirical Software Engineering, 22*(3), 1440–1477. https://doi.org/10.1007/s10664-016-9477-x
177. Johansson, G., Aronsson, G., & Lindström, B.-O. (1978). Social psychological neuroendocrine stress reactions in highly mechanized work. *Ergonomics, 21*(8), 583–599. https://doi.org/10.1080/00140137808931761
178. Korunka, C., Huemer, K.-H., Litschauer, B., Karetta, B., & Kafka-Lützow, A. (1996). Working with new technologies: Hormone excretion as an indicator for sustained arousal. *A pilot study. Biological Psychology, 42*(3), 439–452. https://doi.org/10.1016/0301-0511(95)05172-4
179. Kosti, M. V., Georgiadis, K., Adamos, D. A., Laskaris, N., Spinellis, D., & Angelis, L. (2018). Towards an affordable brain computer interface for the assessment of programmers' mental workload. *International Journal of Human-Computer Studies, 115*, 52–66. https://doi.org/10.1016/j.ijhcs.2018.03.002
180. Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., & Glenk, L. M. (2016). Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in Human Behavior, 62*, 124–135. https://doi.org/10.1016/j.chb.2016.03.081
181. Kuan, K. K. Y., Zhong, Y., & Chau, P. Y. K. (2014). Informational and normative social influence in groupbuying: Evidence from self-reported and EEG data. *Journal of Management Information Systems, 30*(4), 151–178. https://doi.org/10.2753/MIS0742-1222300406
182. Labonté-LeMoyne, É., Santhanam, R., Léger, P.-M., Courtemanche, F., Fredette, M., & Sénécal, S. (2015). The delayed effect of treadmill desk usage on recall and attention. *Computers in Human Behavior, 46*, 1–5. https://doi.org/10.1016/j.chb.2014.12.054
183. Lee, S., Hooshyar, D., Ji, H., Nam, K., & Lim, H. (2018). Mining biometric data to predict programmer expertise and task difficulty. *Cluster Computing, 21*(1), 1097–1107. https://doi.org/10.1007/s10586-017-0746-2

184. Lee, S., Matteson, A., Hooshyar, D., Kim, S., Jung, J., Nam, G., & Lim, H. (2016). Comparing programming language comprehension between novice and expert programmers using EEG analysis. In *Proceedings of the 16th International Conference on Bioinformatics and Bioengineering* (pp. 350–355). https://doi.org/10.1109/BIBE.2016.30
185. Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior, 34*, 273–283. https://doi.org/10.1016/j.chb.2014.02.011
186. Leuthold, S., Schmutz, P., Bargas-Avila, J. A., Tuch, A. N., & Opwis, K. (2011). Vertical versus dynamic menus on the world wide web: Eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference. *Computers in Human Behavior, 27*(1), 459–472. https://doi.org/10.1016/j.chb.2010.09.009
187. Liu, H.-C., Lai, M.-L., & Chuang, H.-H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior, 27*(6), 2410–2417. https://doi.org/10.1016/j.chb.2011.06.012
188. Liu, Y.-F., Kim, J., Wilson, C., & Bedny, M. (2020). Computer code comprehension shares neural resources with formal logical inference in the fronto-parietal network. *eLife, 9,* e59340. https://doi.org/10.7554/eLife.59340
189. Luan, J., Yao, Z., Zhao, F., & Liu, H. (2016). Search product and experience product online reviews: An eye-tracking study on consumers' review search behavior. *Computers in Human Behavior, 65*, 420–430. https://doi.org/10.1016/j.chb.2016.08.037
190. Lundberg, U., Melin, B., Evans, G. W., & Holmberg, L. (1993). Physiological deactivation after two contrasting tasks at a video display terminal: Learning versus repetitive data entry. *Ergonomics, 36*(6), 601–611. https://doi.org/10.1080/00140139308967923
191. Medeiros, J., Couceiro, R., Castelhano, J., Castelo-Branco, M., Duarte, G., Duarte, C., Durães, J., Madeira, H., de Carvalho, P., & Teixeira, C. (2019). Software code complexity assessment using EEG features. In *Proceedings of the 41st International Conference on Engineering in Medicine and Biology Society* (pp. 1413–1416). https://doi.org/10.1109/EMBC.2019.8856283
192. Meppelink, C. S., & Bol, N. (2015). Exploring the role of health literacy on attention to and recall of text-illustrated health information: An eye-tracking study. *Computers in Human Behavior, 48*, 87–93. https://doi.org/10.1016/j.chb.2015.01.027
193. Meservy, T. O., Jensen, M. L., & Fadel, K. J. (2014). Evaluation of competing candidate solutions in electronic networks of practice. *Information Systems Research, 25*(1), 15–34. https://doi.org/10.1287/isre.2013.0502
194. Minas, R. K., Potter, R. F., Dennis, A. R., Bartelt, V., & Bae, S. (2014). Putting on the thinking cap: Using NeuroIS to understand information processing biases in virtual teams. *Journal of Management Information Systems, 30*(4), 49–82. https://doi.org/10.2753/MIS0742-1222300403
195. Molina, A. I., Redondo, M. A., Lacave, C., & Ortega, M. (2014). Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on eye tracking and learner subjective perception. *Computers in Human Behavior, 31*, 475–490. https://doi.org/10.1016/j.chb.2013.04.022
196. Mousavi, S., & Balthazard, P. (2013). Neurophysiologic patterns of learning in decision groups. In *Proceedings of the 46th Hawaii International Conference on System Sciences* (pp. 96–105). https://doi.org/10.1109/HICSS.2013.416
197. Müller, S. C., & Fritz, T. (2016). Using (bio)metrics to predict code quality online. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 452–463). https://doi.org/10.1145/2884781.2884803
198. Nakagawa, T., Kamei, Y., Uwano, H., Monden, A., Matsumoto, K., & German, D. M. (2014). Quantifying programmers' mental workload during program comprehension based on cerebral blood flow measurement: a controlled experiment. In *Proceedings of the Conference Companion of the 36th International Conference on Software Engineering* (pp. 448–451). https://doi.org/10.1145/2591062.2591098

199. Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. (2011). Embodied conversational agent–based kiosk for automated interviewing. *Journal of Management Information Systems, 28*(1), 17–48. https://doi.org/10.2753/MIS0742-1222280102
200. Ortiz de Guinea, A., Titah, R., & Léger, P.-M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. *Journal of Management Information Systems, 30*(4), 179–210. https://doi.org/10.2753/MIS0742-1222300407
201. Pak, J., & Zhou, L. (2013). Eye gazing behaviors in online deception. In *Proceedings of the 19th Americas Conference on Information Systems.*
202. Pantazos, K., & Vatrapu, R. (2016). Enhancing the professional vision of teachers: A physiological study of teaching analytics dashboards of students' repertory grid exercises in business education. In *Proceedings of the 49th Hawaii International Conference on System Sciences* (pp. 41–50). IEEE. https://doi.org/10.1109/HICSS.2016.14
203. Peitek, N., Siegmund, J., Apel, S., Kästner, C., Parnin, C., Bethmann, A., Leich, T., Saake, G., & Brechmann, A. (2020). A look into programmers' heads. *IEEE Transactions on Software Engineering, 46*(4), 442–462. https://doi.org/10.1109/TSE.2018.2863303
204. Pengnate, S. (Fone). (2016). Measuring emotional arousal in clickbait: Eye-tracking approach. In *Proceedings of the 22nd Americas Conference on Information Systems.*
205. Petrusel, R., & Mendling, J. (2013). Eye-tracking the factors of process model comprehension tasks. In C. Salinesi, M. C. Norrie, & Ó. Pastor (Eds.), *Advanced Information Systems Engineering: 25th International Conference, CAiSE 2013, Valencia, Spain, June 17–21, 2013, Proceedings*. LNCS (Vol. 7908, pp. 224–239). Springer. https://doi.org/10.1007/978-3-642-38709-8_15
206. Petrusel, R., Mendling, J., & Reijers, H. A. (2017). How visual cognition influences process model comprehension. *Decision Support Systems, 96*, 1–16. https://doi.org/10.1016/j.dss.2017.01.005
207. Pinggera, J., Furtner, M., Martini, M., Sachse, P., Reiter, K., Zugal, S., & Weber, B. (2013). Investigating the process of process modeling with eye movement analysis. In M. Rosa & P. Soffer (Eds.), *Business Process Management Workshops: BPM 2012 International Workshops, Tallinn, Estonia, September 3, 2012, Revised Papers*. LNBIP (Vol. 132, pp. 438–450). Springer. https://doi.org/10.1007/978-3-642-36285-9_46
208. Randolph, A. B. (2012). Not all created equal: Individual-technology fit of brain-computer interfaces. In *Proceedings of the 45th Hawaii International Conference on System Sciences* (pp. 572–578). https://doi.org/10.1109/HICSS.2012.451
209. Riedl, R., Hubert, M., & Kenning, P. H. (2010). Are there neural gender differences in online trust? An fMRI study on the perceived trustworthiness of eBay offers. *MIS Quarterly, 34*(2), 397–438. https://doi.org/10.2307/20721434
210. Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior, 33,* 201–212. https://doi.org/10.1016/j.chb.2014.01.004
211. Scott, G. G., & Hand, C. J. (2016). Motivation determines Facebook viewing strategy: An eye movement analysis. *Computers in Human Behavior, 56*, 267–280. https://doi.org/10.1016/j.chb.2015.11.029
212. Seo, Y., Kim, M., Jung, Y., & Lee, D. (2017). Avatar face recognition and self-presence. *Computers in Human Behavior, 69*, 120–127. https://doi.org/10.1016/j.chb.2016.12.020
213. Shalom, J. G., Israeli, H., Markovitzky, O., & Lipsitz, J. D. (2015). Social anxiety and physiological arousal during computer mediated vs. face to face communication. *Computers in Human Behavior, 44,* 202–208. https://doi.org/10.1016/j.chb.2014.11.056
214. Shojaeizadeh, M., Djamasbi, S., Chen, P., & Rochford, J. (2017). Task condition and pupillometry. In *Proceedings of the 23rd Americas Conference on Information Systems.*
215. Siegmund, J., Kästner, C., Apel, S., Parnin, C., Bethmann, A., Leich, T., Saake, G., & Brechmann, A. (2014). Understanding source code with functional magnetic resonance imaging.

In *Proceedings of the 36th International Conference on Software Engineering* (pp. 378–389). https://doi.org/10.1145/2568225.2568252
216. Siegmund, J., Peitek, N., Parnin, C., Apel, S., Hofmeister, J., Kästner, C., Begel, A., Bethmann, A., & Brechmann, A. (2017). Measuring neural efficiency of program comprehension. In *Proceedings of the 11th Joint Meeting on Foundations of Software Engineering* (pp. 140–150). https://doi.org/10.1145/3106237.3106268
217. Steinfeld, N. (2016). "I agree to the terms and conditions": (How) Do users read privacy policies online? An eye-tracking experiment. *Computers in Human Behavior, 55*, 992–1000. https://doi.org/10.1016/j.chb.2015.09.038
218. Steinfeld, N., Samuel-Azran, T., & Lev-On, A. (2016). User comments and public opinion: Findings from an eye-tracking experiment. *Computers in Human Behavior, 61*, 63–72. https://doi.org/10.1016/j.chb.2016.03.004
219. Teubner, T., Adam, M. T. P., & Riordan, R. (2015). The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions. *Journal of the Association for Information Systems, 16*(10), 838–879. https://doi.org/10.17705/1jais.00412
220. Trimmel, M., & Huber, R. (1998). After-effects of human-computer interaction indicated by P300 of the event-related brain potential. *Ergonomics, 41*(5), 649–655. https://doi.org/10.1080/001401398186810
221. Tzafilkou, K., & Protogeros, N. (2017). Diagnosing user perception and acceptance using eye tracking in web-based end-user development. *Computers in Human Behavior, 72*, 23–37. https://doi.org/10.1016/j.chb.2017.02.035
222. Astor, P. J., Adam, M. T. P., Jerčić, P., Schaaff, K., & Weinhardt, C. (2013). Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. *Journal of Management Information Systems, 30*(3), 247–278. https://doi.org/10.2753/MIS0742-1222300309
223. Bahr, G. S., & Ford, R. A. (2011). How and why pop-ups don't work: Pop-up prompted eye movements, user affect and decision making. *Computers in Human Behavior, 27*(2), 776–783. https://doi.org/10.1016/j.chb.2010.10.030
224. Bera, P. (2014). Do distracting dashboards matter? Evidence from an eye tracking study. In S. Wrycza (Ed.), *Information Systems: Education, Applications, Research: 7th SIGSAND/PLAIS EuroSymposium 2014, Gdańsk, Poland, September 25, 2014, Proceedings*. LNBIP (Vol. 193, pp. 65–74). Springer. https://doi.org/10.1007/978-3-319-11373-9_6
225. Bera, P. (2016). How colors in business dashboards affect users' decision making. *Communications of the ACM, 59*(4), 50–57. https://doi.org/10.1145/2818993
226. Bina, S., Graue, W., Jones, D. R., Kaskela, T., & Walden, E. A. (2017). Discovery of the optimal visualization for representing three dimensions of data using functional magnetic resonance imaging. In *Proceedings of the 38th International Conference on Information Systems*.
227. Clayton, R. B., Leshner, G., & Almond, A. (2015). The extended iSelf: The impact of iPhone separation on cognition, emotion, and physiology. *Journal of Computer-Mediated Communication, 20*(2), 119–135. https://doi.org/10.1111/jcc4.12109
228. Dianne, C., Head, M., Hector, L., & Bing, P. (2009). Exploring human images in website design: A multi-method approach. *MIS Quarterly, 33*(3), 539–566. https://doi.org/10.2307/20650308
229. Djamasbi, S., Siegel, M., & Tullis, T. S. (2014). Can fixation on main images predict visual appeal of homepages? In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 371–375). https://doi.org/10.1109/HICSS.2014.54
230. Djamasbi, S., Siegel, M., Tullis, T. S., & Dai, R. (2010). Efficiency, trust, and visual appeal: Usability testing through eye tracking. In *Proceedings of the 43rd Hawaii International Conference on System Sciences* (pp. 1–10). https://doi.org/10.1109/HICSS.2010.171
231. Djamasbi, S., Siegel, M., & Tullis, T. S. (2012). Designing noticeable bricklets by tracking users' eye movements. In *Proceedings of the 45th Hawaii International Conference on System Sciences* (pp. 525–532). https://doi.org/10.1109/HICSS.2012.200

232. Djamasbi, S., Siegel, M., & Tullis, T. S. (2012). Faces and viewing behavior: An exploratory investigation. *AIS Transactions on Human-Computer Interaction, 4*(3), 190–211. https://doi.org/10.17705/1thci.00046
233. Durães, J., Madeira, H., Castelhano, J., Duarte, C., & Castelo-Branco, M. (2016). WAP: Understanding the brain at software debugging. In *Proceedings of the 27th International Symposium on Software Reliability Engineering* (pp. 87–92). https://doi.org/10.1109/ISSRE.2016.53
234. Emurian, H. H. (1991). Physiological responses during data retrieval: Comparison of constant and variable system response times. *Computers in Human Behavior, 7*(4), 291–310. https://doi.org/10.1016/0747-5632(91)90017-U
235. Emurian, H. H. (1993). Cardiovascular and electromyograph effects of low and high density work on an interactive information system. *Computers in Human Behavior, 9*(4), 353–370. https://doi.org/10.1016/0747-5632(93)90028-Q
236. Fehrenbacher, D. D., & Smith, S. (2014). Behavioural affect and cognitive effects of time-pressure and justification requirement in software acquisition: Evidence from an eye-tracking experiment. In *Proceedings of the 20th Americas Conference on Information Systems.*
237. Fehrenbacher, D. D., & Tracy, L. M. (2016). Gaze behaviour, motivational factors, and knowledge sharing. In *Proceedings of the 37th International Conference on Information Systems.*
238. Fehrenbacher, D. D. (2017). Affect infusion and detection through faces in computer-mediated knowledge sharing decisions. *Journal of the Association for Information Systems, 18*(10), 703–726. https://doi.org/10.17705/1jais.00470
239. Fehrenbacher, D. D., & Djamasbi, S. (2017). Information systems and task demand: An exploratory pupillometry study of computerized decision making. *Decision Support Systems, 97*, 1–11. https://doi.org/10.1016/j.dss.2017.02.007
240. Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems, 16*(1), 1–47. https://doi.org/10.17705/1jais.00387
241. Girardi, D., Novielli, N., Fucci, D., & Lanubile, F. (2020). Recognizing developers' emotions while programming. In *Proceedings of the 42nd International Conference on Software Engineering* (pp. 666–677). https://doi.org/10.1145/3377811.3380374
242. Gregor, S., Lin, A. C. H., Gedeon, T., Riaz, A., & Zhu, D. (2014). Neuroscience and a nomological network for the understanding and assessment of emotions in information systems research. *Journal of Management Information Systems, 30*(4), 13–48. https://doi.org/10.2753/MIS0742-1222300402
243. Harada, H., Okabe, K., Katsuura, T., & Kikuchi, Y. (1995). Effects of time stress on psychophysiological responses during data entry tasks. *Applied Human Science, 14*(6), 279–285. https://doi.org/10.2114/ahs.14.279
244. Henderson, R., Mahar, D., Saliba, A., Deane, F., & Napier, R. (1998). Electronic monitoring systems: An examination of physiological activity and task performance within a simulated keystroke security and electronic performance monitoring system. *International Journal of Human-Computer Studies, 48*(2), 143–157. https://doi.org/10.1006/ijhc.1997.0167
245. Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in Human Behavior, 50,* 618–625. https://doi.org/10.1016/j.chb.2015.03.017
246. Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology, 92*(1–2), 84–89. https://doi.org/10.1007/s00421-004-1055-z
247. Ikramov, R., Ivanov, V., Masyagin, S., Shakirov, R., Sirazidtinov, I., Succi, G., Thapaliya, A., Tormasov, A., & Zufarova, O. (2019). Initial evaluation of the brain activity under different software development situations. In *Proceedings of the 31st International Conference on Software Engineering and Knowledge Engineering* (pp. 741–747). https://doi.org/10.18293/SEKE2019-215

248. Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior, 32*, 47–53. https://doi.org/10.1016/j.chb.2013.11.013
249. Jay, C., Brown, A., & Harper, S. (2013). Predicting whether users view dynamic content on the world wide web. *ACM Transactions on Computer-Human Interaction, 20*(2), 1–33. https://doi.org/10.1145/2463579.2463580
250. Jenkins, J. L., Anderson, B. B., Vance, A. O., Kirwan, C. B., & Eargle, D. (2016). More harm than good? How messages that interrupt can make us vulnerable. *Information Systems Research, 27*(4), 880–896. https://doi.org/10.1287/isre.2016.0644
251. Johansson, G., & Aronsson, G. (1984). Stress reactions in computerized administrative work. *Journal of Organizational Behavior, 5*(3), 159–181. https://doi.org/10.1002/job.4030050302
252. Kohlisch, O., & Kuhmann, W. (1997). System response time and readiness for task execution the optimum duration of inter-task delays. *Ergonomics, 40*(3), 265–280. https://doi.org/10.1080/001401397188143
253. Kohrs, C., Angenstein, N., Scheich, H., & Brechmann, A. (2012). Human striatum is differentially activated by delayed, omitted, and immediate registering feedback. *Frontiers in Human Neuroscience, 6*, 243. https://doi.org/10.3389/fnhum.2012.00243
254. Kohrs, C., Hrabal, D., Angenstein, N., & Brechmann, A. (2014). Delayed system response times affect immediate physiology and the dynamics of subsequent button press behavior. *Psychophysiology, 51*(11), 1178–1184. https://doi.org/10.1111/psyp.12253
255. Kohrs, C., Angenstein, N., & Brechmann, A. (2016). Delays in human-computer interaction and their effects on brain activity. *PLoS ONE, 11*(1), e0146250. https://doi.org/10.1371/journal.pone.0146250
256. Konok, V., Pogány, Á., & Miklósi, Á. (2017). Mobile attachment: Separation from the mobile phone induces physiological and behavioural stress and attentional bias to separation-related stimuli. *Computers in Human Behavior, 71*, 228–239. https://doi.org/10.1016/j.chb.2017.02.002
257. Kuhmann, W. (1989). Experimental investigation of stress-inducing properties of system response times. *Ergonomics, 32*(3), 271–280. https://doi.org/10.1080/00140138908966087
258. Kuhmann, W., Boucsein, W., Schäfer, F., & Alexander, J. (1987). Experimental investigation of psychophysiological stress-reactions induced by different system response times in human-computer interaction. *Ergonomics, 30*(6), 933–943. https://doi.org/10.1080/00140138708969789
259. Lee, J., & Ahn, J.-H. (2012). Attention to banner ads and their effectiveness: An eye-tracking approach. *International Journal of Electronic Commerce, 17*(1), 119–137. https://doi.org/10.2753/JEC1086-4415170105
260. Li, K., Huang, G., & Bente, G. (2016). The impacts of banner format and animation speed on banner effectiveness: Evidence from eye movements. *Computers in Human Behavior, 54*, 522–530. https://doi.org/10.1016/j.chb.2015.08.056
261. Lo, S.-K., Hsieh, A.-Y., & Chiu, Y.-P. (2014). Keyword advertising is not what you think: Clicking and eye movement behaviors on keyword advertising. *Electronic Commerce Research and Applications, 13*(4), 221–228. https://doi.org/10.1016/j.elerap.2014.04.001
262. Ahrens, M., Schneider, K., & Busch, M. (2019). Attention in software maintenance: An eye tracking study. In *Proceedings of the 6th International Workshop on Eye Movements in Programming* (pp. 2–9). https://doi.org/10.1109/EMIP.2019.00009
263. Mach, Q. H., Hunter, M. D., & Grewal, R. (2010). Neurophysiological correlates in interface design: An HCI perspective. *Computers in Human Behavior, 26*(3), 371–376. https://doi.org/10.1016/j.chb.2009.11.008
264. Moody, G. D., & Galletta, D. F. (2015). Lost in cyberspace: The impact of information scent and time constraints on stress, performance, and attitudes online. *Journal of Management Information Systems, 32*(1), 192–224. https://doi.org/10.1080/07421222.2015.1029391
265. Morana, S., Krönung, J., & Mädche, A. (2017). "I did use it!"—Assessing subjective versus objective cognitive artifact usage. In *Proceedings of the 13th International Conference on Wirtschaftsinformatik* (pp. 1021–1035).

266. Müller, S. C., & Fritz, T. (2015). Stuck and frustrated or in flow and happy: Sensing developers' emotions and progress. In *Proceedings of the 37th International Conference on Software Engineering* (pp. 688–699). https://doi.org/10.1109/ICSE.2015.334
267. Nomura, S., Tanaka, H., & Nagashima, T. (2005). A physiological index of the mental stresses caused by the deskwork with VDT and the relaxation by music. In *Proceedings of the International Conference on Sensors and New Techniques in Pharmaceutical and Biomedical Research* (pp. 130–134). https://doi.org/10.1109/ASENSE.2005.1564522
268. Ortiz de Guinea, A., & Webster, J. (2013). An investigation of information systems use patterns: Technological events as triggers, the effect of time, and consequences for performance. *MIS Quarterly, 37*(4), 1165–1188. https://doi.org/10.25300/MISQ/2013/37.4.08
269. Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117. https://doi.org/10.1016/j.chb.2009.09.001
270. Perrin, J.-L., Paillé, D., & Baccino, T. (2014). Reading tilted: Does the use of tablets impact performance? An oculometric study. *Computers in Human Behavior, 39*, 339–345. https://doi.org/10.1016/j.chb.2014.07.033
271. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2013). Computer breakdown as a stress factor during task completion under time pressure: Identifying gender differences based on skin conductance. *Advances in Human-Computer Interaction, 2013*, 420169. https://doi.org/10.1155/2013/420169
272. Schäfer, F. (1990). The effect of system response times on temporal predictability of work flow in human-computer interaction. *Human Performance, 3*(3), 173–186. https://doi.org/10.1207/s15327043hup0303_3
273. Anderson, B. B., Vance, A. O., Kirwan, C. B., Eargle, D., & Howard, S. (2014). Users aren't (necessarily) lazy: Using NeuroIS to explain habituation to security warnings. In *Proceedings of the 35th International Conference on Information Systems.*
274. Schleifer, L. M., & Okogbaa, O. G. (1990). System response time and method of pay: Cardiovascular stress effects in computer-based tasks. *Ergonomics, 33*(12), 1495–1509. https://doi.org/10.1080/00140139008925349
275. Taylor, B., Dey, A., Siewiorek, D., & Smailagic, A. (2015). Using physiological sensors to detect levels of user frustration induced by system delays. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 517–528). https://doi.org/10.1145/2750858.2805847
276. Thum, M., Boucsein, W., Kuhmann, W., & Ray, W. J. (1995). Standardized task strain and system response times in human-computer interaction. *Ergonomics, 38*(7), 1342–1351. https://doi.org/10.1080/00140139508925192
277. Tozman, T., Magdas, E. S., MacDougall, H. G., & Vollmeyer, R. (2015). Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. *Computers in Human Behavior, 52*, 408–418. https://doi.org/10.1016/j.chb.2015.06.023
278. Trimmel, M., Meixner-Pendleton, M., & Haring, S. (2003). Stress response caused by system response time when searching for information on the Internet. *Human Factors, 45*(4), 615–622. https://doi.org/10.1518/hfes.45.4.615.27084
279. Vance, A. O., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B. (2018). Tuning out security warnings: A longitudinal examination of habituation through fMRI, eye tracking, and field experiments. *MIS Quarterly, 42*(2), 355–380. https://doi.org/10.25300/MISQ/2018/14124
280. Walla, P., & Lozovic, S. (2020). The effect of technology on human social perception: A multi-methods NeuroIS pilot investigation. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience: NeuroIS Retreat 2019*. LNISO (Vol. 32, pp. 63–71). Springer. https://doi.org/10.1007/978-3-030-28144-1_7

281. Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems, 62*, 1–10. https://doi.org/10.1016/j.dss.2014.02.007
282. Wang, Q., Xu, Z., Cui, X., Wang, L., & Ouyang, C. (2017). Does a big Duchenne smile really matter on e-commerce websites? An eye-tracking study in China. *Electronic Commerce Research, 17*(4), 609–626. https://doi.org/10.1007/s10660-016-9237-4
283. Wu, Y.-L., & Chen, P.-C. (2016). The synesthesia effects of online advertising stimulus design on word-of-mouth and purchase intention: From the perspective of consumer olfactory and gustatory. In *Proceedings of the 22nd Americas Conference on Information Systems.*
284. Anderson, B. B., Vance, A. O., Kirwan, C. B., Eargle, D., & Jenkins, J. L. (2016). How users perceive and respond to security messages: A NeuroIS research agenda and empirical study. *European Journal of Information Systems, 25*(4), 364–390. https://doi.org/10.1057/ejis.2015.21
285. Züger, M., & Fritz, T. (2015). Interruptibility of software developers and its prediction using psycho-physiological sensors. In *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems* (pp. 2981–2990). https://doi.org/10.1145/2702123.2702593
286. Züger, M., Müller, S. C., Meyer, A. N., & Fritz, T. (2018). Sensing interruptibility in the office: A field study on the use of biometric and computer interaction sensors. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1–14). https://doi.org/10.1145/3173574.3174165
287. Anderson, B. B., Vance, A. O., Kirwan, C. B., Jenkins, J. L., & Eargle, D. (2016). From warning to wallpaper: Why the brain habituates to security warnings and what can be done about it. *Journal of Management Information Systems, 33*(3), 713–743. https://doi.org/10.1080/07421222.2016.1243947
288. Anderson, B. B., Jenkins, J. L., Vance, A. O., Kirwan, C. B., & Eargle, D. (2016). Your memory is working against you: How eye tracking and memory explain habituation to security warnings. *Decision Support Systems, 92*, 3–13. https://doi.org/10.1016/j.dss.2016.09.010
289. Loos, P., Riedl, R., Müller-Putz, G. R., vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering, 2*(6), 395–401. https://doi.org/10.1007/s12599-010-0130-8
290. Riedl, R., Davis, F. D., Banker, R. D., & Kenning, P. H. (2017). *Neuroscience in information systems research: Applying knowledge of brain functionality without neuroscience tools.* Springer. https://doi.org/10.1007/978-3-319-48755-7
291. Senior, C., Russell, T., & Gazzaniga, M. S. (2009). *Methods in mind*. MIT Press.
292. Rohleder, N., Wolf, J. M., Maldonado, E. F., & Kirschbaum, C. (2006). The psychosocial stress-induced increase in salivary alpha-amylase is independent of saliva flow rate. *Psychophysiology, 43*(6), 645–652. https://doi.org/10.1111/j.1469-8986.2006.00457.x
293. Chen, Y., Fang, W., Guo, B., & Bao, H. (2021). Fatigue-related effects in the process of task interruption on working memory. *Frontiers in Human Neuroscience, 15*, 703422. https://doi.org/10.3389/fnhum.2021.703422
294. Stangl, F. J., Riedl, R., Kiemeswenger, R., & Montag, C. (2023). Negative psychological and physiological effects of social networking site use: The example of Facebook. *Frontiers in Psychology, 14*, 1141663. https://doi.org/10.3389/fpsyg.2023.1141663
295. Tarafdar, M., Gupta, A., & Turel, O. (2013). The dark side of information technology use. *Information Systems Journal, 23*(3), 269–275. https://doi.org/10.1111/isj.12015

Investigating the Impact of Mood and Emotion on the Perception of Fake News on Social Media Platforms

Rana Ali Adeeb and Mahdi Mirhoseini

Abstract Social media platforms have enabled producers of fake news to craft messages at an unprecedented rate and have rendered these messages accessible to the masses at no cost. Although the context of social media is ripe with affective attributes, research on fake news on social media has largely strayed away from understanding the impact of users' emotion on their perception of fake news. We propose a NeuroIS approach to investigate the effect of experiencing specific moods and emotions on (a) users' belief in fake news and (b) users' intent to share misinformation in the context of users interacting with social media. This research-in-progress responds to calls for research on the mechanisms by which fake news become entrenched as well as mechanisms to quell the influence of fake news. It also contributes to the emerging literature on the role of emotion in the perception of fake news on social media platforms.

Keywords Fake news · Emotion · Belief · Intent to share · EDA · Face reader

1 Introduction

Although news fabrication has been present in the media since the early twentieth century [15], social media platforms that have emerged within the scope of new technologies allow producers of fake news to craft messages at an unprecedented scale and provide novel ways in which problematic information is created, disseminated, and perceived by users [17]. These platforms are ripe with emotionally charged pieces of information that users are exposed to as they browse their social media platforms. This can induce different emotions which may impact social media users' subsequent interaction with news related posts leading to different types of behavior

R. Ali Adeeb (✉) · M. Mirhoseini
John Molson School of Business, Concordia University, Montreal, Canada
e-mail: rana.aliadeeb@mail.concordia.ca

M. Mirhoseini
e-mail: mahdi.mirhoseini@concordia.ca

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_12

such as sharing, commenting, messaging, and liking [9]. The absence of affect in understanding users' belief in misinformation and informing social media design may be one of the reasons for the limited success of existing intervention methods in improving users' ability to discern fake news from real news [15]. Consequently, understanding the role affect plays in the perception of fake news on social media may provide insight into the mechanisms by which fake news become entrenched as well as the mechanisms to counter the flow and influence of fake news on social media platforms. Yet, most studies focusing on combatting fake news on social media have focused on the impact of cognitive factors such as analytical thinking [22, 23], dogmatism [5], and fact-checkers [15] and have largely strayed away from understanding the impact of emotion on the perception of fake news. This is surprising because a salient feature of fake news headlines on social media platforms is that they are often emotionally evocative [25]. Even more surprising is the fact that the impact of emotion on human judgement and decision making is well documented in the Psychology literature [1] by the dual-process theories of cognition [12], the assimilative accommodative model of emotion [3], and feelings as information theory [27]. One reason for the scant literature on the impact of emotion on the perception of fake news is that emotion is a variable that is difficult to study because it is prompted through implicit bodily processes outside one's conscious awareness [6]. Additionally, emotion evolves quickly as users interact with the abundantly available and emotionally charged elements on social media platforms.

To address this gap, we propose a NeuroIS study to understand the role of two affective variables: (i) mood and (ii) emotion in influencing (a) users' belief in fake news on social media platforms and (b) users' intent to share fake news on social media platforms. We deploy the neurophysiological measure of emotion of Electrodermal activity (EDA), which captures with high temporal resolution [6] how emotion evolves. We also use Face reader technology, which captures changes in users' facial expression as a result of experiencing emotions [8].

2 Literature Review

The fake news phenomenon, characterized as "fabricated information that mimics news media content in form but not in organizational process or intent" [15, p. 1094], has raised concerns about the potential threat posed by the Internet and social media at individual and societal levels [5]. Because social media platforms offer a means for communication through which emotional content is propagated, "news media coverage and user interactions on social media frequently exhibit a shrill emotional component" [28, p. 232]. By considering the emotional factors that affect users' responses to fake news, future interventions may have a better chance of success in combatting this societal problem. Core affect is a neurophysiological state consciously accessible as a non-reflective feeling that is an integral blend of hedonic or valence value (pleasure–displeasure) and arousal or activation value (sleepy–activated) [26]. We use the Affective Response Model (ARM) [30] conceptual framework to classify two groups of affective variables and investigate their effect on users' belief in fake news and users' intent to share misinformation in the context of users interacting with social media: (a) Mood: a prolonged core affect that has an unclear or unknown stimulus [26] and (b) Emotion: an affective state induced as a result of an interaction between users and a stimulus [26]. In investigating the relationship between affect and individuals' propensity to believe information, two competing theories exist. The assimilative-accommodative model [3] states that positive and negative emotions influence peoples' perceived accuracy of information by regulating their information processing strategies differently. Specifically, individuals experiencing positive emotion tend to employ more heuristic strategies while those in negative emotional states tend to use more effortful processing strategies [3]. The fake news literature in support of this theory is scarce. The resource allocation model [10], which can be classified under the dual process models of cognition [29] on the other hand states that both positive and negative emotions facilitate heuristic information processing strategies because they increase irrelevant thoughts that occupy attentional resources and reduce the processing effort invested in cognitive tasks [11]. This theory was informed by Martel et al. [18] who investigated the role of emotion on the likelihood of believing fake news. In one study, they reported a positive correlation between self-reported use of emotion and belief in fake (but not real) news and found that the more participants depended on emotion over reason, the more they perceived fake stories as accurate. In a separate study, [18] demonstrated that an increased reliance on momentary emotion, regardless of type or valence, increased susceptibility to fake (but not real) news on social media above and beyond a simple lack of reasoning and decreased discernment between real and fake news. This finding was also demonstrated by Rosenzweig et al. [25] who found that experiencing any emotional reaction was associated with worse truth discernment of headlines and that respondents were better at discerning true from false news when they experienced no emotion after reading a headline. These studies combined suggest that affect may play a unique role in users' susceptibility to fake news on social media platforms.

On the sharing of content on social media, the veracity of headlines was found to have little effect on sharing intentions, despite significantly influencing accuracy judgements [20, 21]. This accuracy-sharing dissociation phenomenon [20, 21] can be explained by three competing theories: (1) the confusion-based account, suggesting that people mistakenly believe false claims they share are true; (2) the preference-based account, indicating that people prioritize political identity over truth and share consistent but false content; and (3) the inattention-based account, proposing that social media distracts people from their preference for sharing only accurate content.

3 Research Model

We aim to understand how mood and experiencing different emotions impact users' belief and truth discernment and users' intent to share fake news on social media.

Impact of Mood on Belief in and Intent to Share Fake News

In line with the resource allocation model [10] and findings by Martel et al. [18] and Rosenzweig et al. [25], we predict that heightened mood, regardless of valence, increases the extent to which people believe fake news on social media. Thus:

H1: Experiencing high levels of positive or negative mood is associated with belief in fake news headlines on social media platforms.

In line with the confusion-based account [20, 21] we predict that heightened mood increases the extent to which people intend to share fake news headlines. Hence:

H2: Experiencing high levels of positive or negative mood is associated with an increased intent to share fake news headlines on social media platforms.

Impact of Emotion on Belief in and Intent to Share Fake News

As users browse social media, they go on an emotion rollercoaster which can impact their perception of and behavior regarding news related content they come across. In line with the resource allocation model [10], we predict that heightened emotionality, regardless of valence, increases the extent to which people believe fake news. Thus:

H3: Experiencing high levels of emotion is associated with belief in fake news headlines on social media platforms.

We predict, in line with the confusion-based account [20, 21] that heightened emotion increases the extent to which people intend to share fake news headlines. Hence:

H4: Experiencing high levels of emotion is associated with an increased intent to share fake news headlines on social media platforms.

4 Methodology

Mood Task

We investigate the impact of state-based emotionality (users' mood) on (a) users' accuracy judgements of and (b) users' intent to share real and fake headlines on social media. We control for a range of individual difference factors such as analytical

Fig. 1 Example of headline (left) used in mood task and meme (right) used in emotion task

thinking and dogmatism. To test H1 and H2, eight headlines that mimic Facebook posts were chosen according to the bipartisan factchecker website snopes.com, were adjusted for their length in terms of the number of words, and were balanced in terms of valence, truthfulness, and political stance. A control headline, the assessment of which is very easy was also created. Figure 1 shows an example of one of the headlines. The nine headlines will be randomly presented to participants who will be asked to assess the veracity of each headline and then report on their intent to share it.

Emotion Task

Next, we investigate the impact of induced emotionality (users' emotion) on (a) users' accuracy judgements of and (b) users' intent to share real and fake headlines on social media. We control for the same individual difference factors in the mood task. A one factor (valence) within subject experiment will test H3 and H4. Ten memes have been created and balanced in terms of valence. Two news headlines were chosen according to the bipartisan factchecker website snopes.com, were adjusted for their length in terms of the number of words and were balanced in terms of valence. Figure 1 shows an example of one of the memes. Two blocs (five negatively valenced memes followed by a target headline and five positively valenced memes followed by a target headline) will be randomly presented to participants who will be asked to assess the veracity of each target headline and subsequently report on their intent to share it.

Pinpointing the Accuracy-Sharing Dissociation

In both the mood and the emotion tasks, participants will be asked to rate the accuracy of each headline immediately before making a judgement about sharing the headline (i.e., a full-attention treatment [20]). Thus, we will be able to distinguish between false items that: (i) participants share and also believe to be accurate (confusion-based rejection of truth); (ii) participants share despite believing to be inaccurate (preference-based rejection of truth); and (iii) participants did not share once they considered accuracy (inattention-based).

Other Potential Impacts of Affective Variables

The design of the present research allows us to further study other potential impacts of affective variables and their interaction on belief and behavior. Because our experiment involves two tasks (the mood task followed by the emotion task), we would be able to investigate whether, as a result of browsing and experiencing different emotional states, the initial impact of mood on belief and intent to share is reduced and replaced by momentary emotion. Additionally, the present research would allow us to examine whether positive or negative emotion has a differential impact on belief as modeled by the assimilative model.

Measurement

Participants' momentary mood state will be measured using the self-perceived scale PANAS [31] and circumplex model of emotion [19, 24] prior to engaging with stimuli. The circumplex model is based on two independent dimensions: valence (positive–negative continuum) and arousal (deactivated–activated continuum). Participants' emotion will be measured using neurophysiological measures prior to engaging with the stimuli in the emotion task. The valence dimension will be measured using Face reader technology, which captures changes in users' facial expression as a result of experiencing emotions. We use Facereader technology (Noldus Information Technology, Wageningen, The Netherlands) which classifies users' facial expressions and generates a valence score [8]. The arousal dimension will be captured using the Electrodermal activity (EDA) technique, a biomarker of arousal with well-known psychophysiological functioning [7] and which provides insight into the automatic affective processes. EDA captures skin connectivity by attaching two electrodes to users' fingers [6]. Using a neurophysiological measure of emotion is of value because bodily responses constitute a critical element of the emotional experience overall [14]. Additionally, since emotion is a very fast process that starts within milliseconds of interacting with a stimulus, we chose the EDA neurophysiological

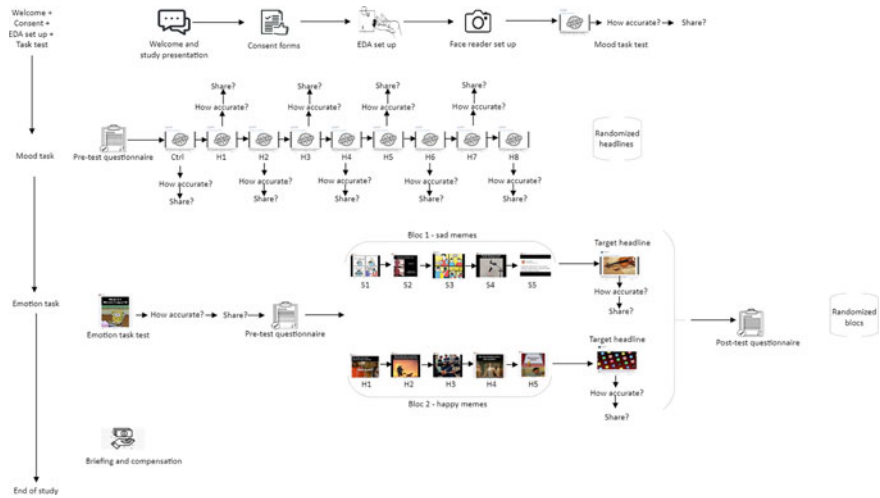


Fig. 2 Experimental design

measure of emotion because it is capable of capturing users' affective reactions with millisecond precision [6]. Our approach of integrating the ARM model, with the two existing theories on emotion (i.e., The assimilative-accommodative and the resource allocation models) and using neurophysiological data as complementary sources of information to traditional self-reported measures provides a unique opportunity to understand if and how users' affective state impacts their interaction with fake news on social media. Figure 2 demonstrates the experimental design of the present research.

Apparatus and Procedure

Participants will be welcomed to the test site and their informed consent form will be obtained. The experiment will be carried out in a room set to a temperature of 25–26 C with 50% humidity [6]. We use a Logitech C600 webcam mounted on a laptop facing the participants and FaceReader4 software [8]. Participants will be instructed to look directly towards the camera while showing their facial expression. Recordings with a resolution of 640 × 480 at 25 frames per second will be analyzed frame by frame, scaling the six basic emotions and neutral from 0 (not present at all) to 1 (maximum intensity of the fitted model). To set up the EDA measure, electrodes will be placed on the distal phalanges of the pointing and middle finger of the participants' non-dominant hand [6]. To account for the physiological latency of the skin conductance rate (SCR), a window between 1 and 4- or 5-s following stimulus presentation will be used to identify the time window that the SCR rise is assumed to happen [16]. ISCR, an integrated SCR measure [2] will be used to obtain an unbiased estimate of the total SCR magnitude within the response window. We will introduce some variability "jittering" to reduce motor and psychological preparation that participants could—consciously or not—establish as they go through the experiment [6]. The data will be down-sampled to 10 Hz from the original 500 or 1000 Hz and a low pass filter (5 Hz, first-order Butterworth) will be applied. The minimum threshold amplitude will be set at 0.01 micro-Siemens to remove noise [4].

5 Expected Contributions

The present research will contribute to the fake news literature through its focus on existing gaps in our understanding of the role of affective variables in the perception of misinformation on social media platforms. It responds to calls for research on the mechanisms by which fake news become entrenched as well as mechanisms to quell the influence of fake news on social media. This is an important topic because an increasing number of people are obtaining their news via social media platforms which rely on unknown sources [13]. This research would also contribute to the emerging literature on the role of emotion in the perception of fake news. Understanding the role that emotion plays in the belief and intent to share fake news on social media platforms could offer a general framework that can help explain various findings about the perception of fake news on social media and can inform the design of IT based intervention methods such as fact checkers to counter the proliferation of fake news that we are witnessing today. Consequently, the present research can have implications for technology platforms, governments, and citizens interested in combating infodemics.

References

1. Ajzen, I., Fishbein, M., Lohmann, S., & Albarracín, D. (2018). The influence of attitudes on behavior. In *The handbook of attitudes* (pp. 197–255).
2. Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*, 80–91.
3. Bless, H., Fiedler, K., & Forgas, J. P. (2006). Mood and the regulation of information processing and behavior. In *Affect in social thinking and behavio*r (pp. 65–84).
4. Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.
5. Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition, 8*, 108–117.
6. Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The body and the brain: Measuring skin conductance responses to understand the emotional experience. *Organizational Research Methods, 22*, 394–420.
7. Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *Journal of Neuroscience, 20*, 3033–3040.

8. Danner, L., Sidorkina, L., Joechl, M., & Duerrschmid, K. (2014). Make a face! Implicit and explicit measurement of facial expressions elicited by orange juices using face reading technology. *Food Quality and Preference, 32*, 167–172.
9. Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science, 31*, 75–87.
10. Ellis, H. C., & Ashbrook, P. W. (1989). The "state" of mood and memory research: A selective review. *Journal of Social Behavior and Personality, 4*(2), 1.
11. Ellis, H. C., Ottaway, S. A., Varner, L. J., Becker, A. S., & Moore, B. A. (1997). Emotion, motivation, and text comprehension: The detection of contradictions in passages. *Journal of Experimental Psychology: General, 126*(2), 131–146.
12. Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223–241.
13. Gottfried, J., & Shearer, E. (2016). *News use across social media platforms 2016*.
14. Grandjean, D., Sander, D., & Scherer, K. R. (2008). Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and Cognition, 17*, 484–495.
15. Lazer, D. M. J. (2018). The science of fake news. *Science, 359*, 1094–1096.
16. Levinson, D. F., & Edelberg, R. (1985). Scoring criteria for response latency and habituation in electrodermal research: A critique. *Psychophysiology, 22*, 417–426.
17. Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6*, 353–369.
18. Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research, 5*, 47.
19. Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.
20. Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature, 592*(7855), 590–595.
21. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770–780.
22. Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowd-sourced judgments of news source quality. *Proceedings of the National academy of Sciences of the United States of America, 116*, 2521–2526.
23. Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition, 188*, 39–50.
24. Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Developmental Psychopathology, 17*.
25. Rosenzweig, L. R., Bago, B., Berinsky, A. J., & Rand, D. G. (2021). Happiness and surprise are associated with worse truth discernment of COVID-19 headlines among social media users in Nigeria. *HKS Misinfo Review*.
26. Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.
27. Schwarz, N. (2012). Feelings-as-information theory. In *Handbook of theories of social psychology* (Vol. 1, pp. 289–308). SAGE Publications Ltd.
28. Serrano-Puche, J. (2021). Digital disinformation and emotions: Exploring the social risks of affective polarization. *International Review of Sociology, 31*, 231–245.
29. Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.
30. Syracuse University, & Zhang, P. (2013). The affective response model: A theoretical framework of affective concepts and their relationships in the ICT context. *MISQ, 37*, 247–274.

31. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.

Connecting Values: The Research Potential of Sustainable Engagement

Elaine Mosconi, Marie-Claude Boudreau, and Leandro Feitosa Jorge

Abstract Organizational sustainability has the potential to connect organizational values with those of their employees, fostering their engagement. Employee engagement is a central concern in today's business environment, as it is related to increased employee productivity, well-being, loyalty to the organization, and higher productivity. While many studies focus on the characteristics and effects of social media in organizations, few studies analyze how the type of content posted affects employee engagement. Hence, this research-in-progress describes a systematic literature review to understand how previous research analyzed the relationship between enterprise social media and employee engagement. More specifically, this research aims to outline relevant applications for using neurophysiology to study the relationship between organizational social media posts that highlight their sustainable values and employee engagement and to instigate further research in the study domain, both theoretical and empirical.

Keywords Employees · Engagement · Sustainable values · Enterprise social media · Neuroscience

1 Introduction

The digital business environment describes the combination of different digital business ecosystems integrated into a sociotechnical network of individuals, organizations, and technologies that work collectively to co-create value [1]. The digital

E. Mosconi (✉) · L. F. Jorge
Business School, Université de Sherbrooke, Sherbrooke, Canada
e-mail: elaine.mosconi@usherbrooke.ca

L. F. Jorge
e-mail: leandro.feitosa.jorge@usherbrooke.ca

M.-C. Boudreau
Terry College of Business, University of Georgia, Athens, USA
e-mail: mcboudre@terry.uga.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_13

business environment brings many benefits as working-from-home conditions but also imposes high speed and an "always connected" work conditions, requiring more meaningful connections between employees and organizational values. In the digital business environment, the use of social media has spread to the organizational context, affecting many organizational activities that go beyond sales and customer relationships [2]. One of the most relevant and studied effects of using social media within organizations is employee engagement [3]. However, most research at the intersection of social media and employee engagement focuses on the behaviors and practices afforded by this type of technology, describing the use of company social media to give employees a voice [4], generate new ideas [5], and empower the workforce [6].

Employee engagement is a positive attitude of employees towards their organization and its values [7]. It relates to superior organizational performance outcomes, such as employee retention, productivity, profitability, customer loyalty, and safety [8]. Hence, combining these positive outcomes with the network potential of social media seems to be strategic for organizations, as employees can act as brand ambassadors internally and externally [9]. However, when employees' values are disconnected from organizational values, the effects on the organizational brand can be negative [10].

In the past decades, we have observed a shift in organizational values leading to a greater embrace of the triple bottom line of sustainability (i.e., planet, people, and prosperity) and an expectation for firms to report to behave sustainably. Accordingly, there has been a steady increase in corporate social responsibility, with many firms integrating environmental and social performance criteria in executive compensation [11]. One of the most respected sustainability certifications is the B Corp certification. Currently, over 6,000 organizations worldwide have received this certification. B Corps are a model of companies that meet the highest standards of social and environmental performance, public transparency, and corporate responsibility to balance profit and purpose [11]. As their tagline states, B Corps constitute a "force for good". The B Corp certification process has been related to increased stakeholder engagement and brand reputation [12]. As well, recent research has shown that the B Corp certification helps ingrain an organization's mission in its employees' mindset and thus supports recruitment and employee engagement, which in turn contributes to internal alignment [13].

Against the background, we outline relevant application areas for using biosignals and neurophysiology in connecting enterprise social media (via posts related to an organization's sustainability actions) and employee engagement in the context of B Corp. We believe B Corp organizations constitute a particularly appropriate setting, given these organizations' demonstrated commitment towards social and environmental matters. In addition, following the IS research stream on sustainability, we posit that using the Neuro IS application and bio-signals of employees within enterprise social media use and interaction regarding sustainable practices can help to understand their behavior and improve their engagement when it resonates with their personal values. However, Information systems (IS) literature falls short in analyzing the relationship between employees' posting content related to the triple bottom line of sustainability (i.e., planet, people, and prosperity) on the internal platforms and their engagement with the company.

Hence, this research presents a systematic literature review of enterprise social media and employee engagement to identify future research gaps that could include Neuro IS strategies considering that neurophysiology and biosignal are able to provide accurate data to work on employee engagement towards sustainability objectives. Then, this paper aims to instigate further research on this topic, both theoretical and empirical. The findings of this systematic literature review contribute to improving our understanding of the impact of employees' internal posting on sustainability-related content and their engagement, as well as the role of enterprise social media in facilitating this process. By addressing this research gap, the study will offer valuable insights for organizations seeking to enhance their sustainability efforts and promote employee engagement.

2 Systematic Literature Review Methodology

Rousseau et al. [12, p. 7] argue that literature reviews are a "comprehensive accumulation, transparent analysis, and reflective interpretation of all empirical studies about a specific question" [14]. Following this perspective, we adopted the guidelines proposed by Okoli and Schabram [13] to analyze previous empirical studies on our research question [15]: "How does enterprise social media support employees' engagement?"

To conduct the review, we used the following criteria for text extraction: (1) objective: understand enterprise social media value-creation in the digital business environment; (2) databases: SCOPUS, and SCIENCE DIRECT, (3) search: empirical peer-reviewed academic journal articles; (4) practical screen: journal articles describing the use of enterprise social media and its effects on employee engagement; (5) quality appraisal: meet the established quality criteria, we restricted our search only to academic journals ranked A *, A, and B in the Australian Business Dean's Council journal ranking scheme and/or 4 ∗ , 4, and 3 in the Chartered Association of Business Schools journal ranking scheme, as the two journal ranking schemes were used because of their wide acceptance in the business research community. To extract the data (step 6), we screened the selected papers focusing on identifying *how* enterprise social media supported employees' engagement. The adoption of the internal perspective excluded the use of enterprise social media for brand management, customer relations, marketing, and sales from the analyses.

We applied a search on "TITLE" and "KEYWORDS:" and selected the filters: "English" for the language of the papers and "peer-reviewed" and "academic journals." We included only academic journals in our sample as part of the "quality appraisal" criteria of the systematic literature review protocol. The selection criteria resulted in 202 papers, with no duplication between the two databases. Based on the quality criteria and the focus of this research, the final sample yielded 10 journal articles, as described in Fig. 1.

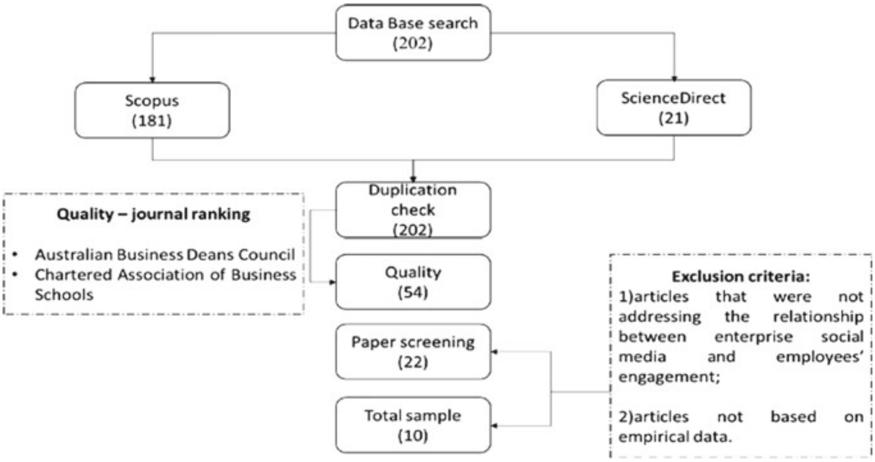


Fig. 1 Papers selection process

3 Results

Results of the systematic literature review show that Enterprise social media (ESM) can have a significant impact on employee engagement. The analyzed sample describes that effective use of ESM can improve communication, work-life balance, and leadership in organizations [16]. The organizational socialization that occurs in virtual settings, such as ESM, is also unique and can impact employee engagement, helping employees feel more connected to the organization and improving engagement over time. The relationship between ESM use and employee engagement can also be influenced by factors such as organizational values, leadership profiles, and psychological conditions. ESM can also facilitate organizational socialization and help employees navigate the workplace of the future. By facilitating digital work, organizations can improve employee engagement and productivity [17].

Overall, ESM can positively and negatively affect employee engagement, depending on how it is used and managed. Organizations should strive to create a culture that supports effective ESM use, provides guidelines and training for employees, and encourages work-related use of ESM. The papers' sample explore the relationship between enterprise social media (ESM) usage and employee engagement. Table 1 summarizes the key findings from each article and their main implications for organizations and employees. Furthermore, fostering open communication and collaboration through ESM platforms can also positively impact employee engagement and contribute to a more engaged and connected workforce.

The papers' sample highlights the need for further research on ESM and its impact on employee engagement, particularly in diverse industries and organizational contexts.

As shown in the results of the systematic literature review, the alignment of employees' sustainability values with organizational values is a gap in the current

Table 1 Key findings from the papers' sample

Reference	Key findings
Cai et al. [16, 18]	ESM usage positively affects psychological conditions such as job satisfaction, organizational commitment, and self-efficacy, which in turn enhances employee agility
Dittes et al. [17, 19]	ESM use is connected to the workplace of the future, and its use can support employee engagement by providing opportunities for feedback, recognition, and collaboration
Ewing et al. [20, 21]	One relevant factor in promoting employees' engagement through ESM seems to be the alignment between employees' values and organizational values
Gupta et al. [22]	Socialization in a virtual setting can provide significant support to new employees. It is more effective when combined with in-person interactions
Huang et al. [23]	Propose the concept of communicational ambidexterity, which refers to the ability to effectively manage both formal organizational content with user-generated content. The study emphasizes the need for organizations to support multiple voices without drowning out a unified corporate message
Leidner et al. [24]	ESM affordances can facilitate the socialization of new employees and promote employee engagement. ESM can help create a sense of community and belonging among employees, which in turn, can contribute to higher levels of job satisfaction and engagement
Men et al. [25]	The more employees read organizations' and co-workers' posts or interact with them (liking or commenting), the more they feel engaged with the organizations. This engagement is related to perceived organizational transparency and employee identification with the organization
Nivedhitha and Manzoor [26]	Examine the impact of ESM on reducing cyberslacking and increasing employee engagement, highlighting a significant mediating effect of workplace social bonding
Oksa et al. [27]	The study shows that professional social media invasion can positively affect work engagement, which may be explained by how ESM supports employees' ability to manage and combine their work and private boundaries effectively. The study also highlights that no association between work exhaustion and professional social media invasion-enabled productivity was established
Song et al. [14]	Research results stress that work-oriented social media and socialization-oriented social media are complementary resources that generate synergies to improve team and employee performance. Activities enabled by socialization-oriented social media enhance employee attachment and belonging, positively affecting employees' commitment to the team and organization

literature. However, it is an important factor in promoting employee engagement [21]. Hence, this work-in-progress proposes the analysis of biosignals to study the physiological responses of employees as they use social media to share their organizations' sustainable actions with other employees.

4 Discussion and Further Research

The use of biosignals to evaluate employees' engagement in social media is a relatively new area of research, and there is limited scientific literature on the topic. However, some studies have investigated physiological measures to assess emotional responses to social media content, which could provide the basis for this study. For instance, Deitz et al. [15] examined the relationship between electroencephalogram (EEG) signals and social media engagement. The study aimed to provide insights into the effectiveness of Super Bowl ads on YouTube by comparing two different measures of consumer behavioral responses: self-report USA Today Ad Meter ratings and EEG-based neural engagement scores. Another study by Kosch et al. [16] explored facial expression analysis to measure emotional responses to smartphone use. Their results described an accessible technology for emotion recognition by assessing its feasibility in natural and uncontrolled setups. Similarly, a study by Rúa-Hidalgo et al. [17] investigated the use of physiological measures to assess emotional responses to social media gifts on Instagram. The researchers used a combination of EEG and galvanic skin response (GSR) to measure cognitive and emotional engagement.

While these studies suggest that biological signals can be used to assess emotional responses to social media and mobile content, more research is needed to determine how these measures can precisely assess the relationship between employee engagement and organizational values. In this context, this research suggests monitoring physiological signals, such as heart rate, skin conductance, and brain waves, when employees post and interact with content linked to the scope of Bcorp certification. These signals can be measured using wearable devices such as smartwatches or fitness trackers, which can be worn by employees while they are using enterprise social media.

To achieve this, we plan:

(1) to establish a baseline for each employee by measuring their biosignals during a neutral task, such as reading a book or watching a video. This baseline can then be used to compare their biosignals during social media use.
(2) during social media use, employees can be presented with content that relates to the organization's values, and their biosignals can be monitored to evaluate their engagement with this content. For example, if an employee is presented with content related to sustainability and their heart rate increases, this may indicate that they are highly engaged with the content and the organization's sustainability values.
(3) we can also investigate employee engagement at different levels by using different experiential designs. For example, we can complement the data collected using bio-signals with employees' perceptions after the experience with focus groups, interviews, and surveys, to measure overall employee engagement, as well as the engagement with specific aspects of the certification process.

We build on previous research that have measure already individuals' engagement on social media … This data can be used to observe how employees react to different types of posts and to identify which posts are more engaging to them. In addition, by using NeuroIS software, we can investigate the differences in engagement between different types of employees, such as new hires versus senior employees.

5 Conclusion

One of the challenges in the digital and globalized business environment that we are facing today is attracting and retaining talent. Considering the new generation's profile, matching organizational values with employees' values seems to be a pivotal path to accomplish this. Hence, in this work-in-progress paper, we propose to integrate Neuro IS strategies to analyze the alignment between employees and organizational values using enterprise social media platforms, especially in contexts where companies are pursuing a journey for B Corp certification. In addition, this paper also highlights the sustainable dimension of digital transformation, which is often neglected, as most studies focus on economic benefits rather than social and environmental ones. It includes investigating the neural pathways that lead to increased engagement and performance, as well as the physiological and psychological responses associated with engagement with organizational values, and gaining insights into how organizational values and processes impact employee engagement and performance.

References

1. Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International Journal of Information Management, 47*, 52–64.
2. Jorge, L. F., Mosconi, E., & Santa-Eulalia, L. A. (2022). Enterprise social media platforms for coping with an accelerated digital transformation. *Journal of Systems and Information Technology*.
3. Sharma, A., & Bhatnagar, J. (2016). Enterprise social media at work: Web-based solutions for employee engagement. *Human Resource Management International Digest*.
4. Estell, P., & Davidson, E. J. (2019). *Employee engagement, voice mechanisms, and enterprise social network sites (ESNS)*.
5. Gode, H. E., Johansen, W., & Thomsen, C. (2019). Employee engagement in generating ideas on internal social media: A matter of meaningfulness, safety and availability. *Corporate Communications: An International Journal, 25*, 263–280.
6. Haddud, A., Dugger, J. C., & Gill, P. (2016). Exploring the impact of internal social media usage on employee engagement. *Journal of Social Media for ORGANIZATIONS, 3*, 1–23.
7. Robinson, D., Perryman, S., & Hayday, S. (2004). *The drivers of employee engagement*. Institute for Employment Studies.
8. Markos, S., & Sridevi, M. S. (2010). Employee engagement: The key to improving performance. *International Journal of Business and Management, 5*, 89.
9. Gelb, B. D., & Rangarajan, D. (2014). Employee contributions to brand equity. *California Management Review, 56*, 95–112.

10. Mazzei, A., & Quaratino, L. (2017). The influence of the organizational and managerial context on employee internal voice and brand ambassador behaviour: A study of a retail company. *Journal of Management & Change*.
11. B Lab Global Site. https://www.bcorporation.net/en-us
12. Rousseau, D. M., Manning, J., & Denyer, D. (2008). 11 evidence in management and organizational science: Assembling the field's full weight of scientific knowledge through syntheses. *Academy of Management Annals, 2*, 475–515.
13. Okoli, C., & Schabram, K. (2010). *A guide to conducting a systematic literature review of information systems research*.
14. Song, Q., Wang, Y., Chen, Y., Benitez, J., & Hu, J. (2019). Impact of the usage of social media in the workplace on team and employee performance. *Information & Management, 56*, 103160.
15. Deitz, G. D., Royne, M. B., Peasley, M. C., & Coleman, J. T. (2016). EEG-based measures versus panel ratings: Predicting social media-based behavioral response to Super Bowl ads. *Journal of Advertising Research, 56*, 217–227.
16. Kosch, T., Hassib, M., Reutter, R., & Alt, F. (2020). Emotions on the go: Mobile emotion assessment in real-time using facial expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 1–9).
17. Rúa-Hidalgo, I., Galmes-Cerezo, M., Cristofol-Rodríguez, C., & Aliagas, I. (2021). Understanding the emotional impact of gifs on Instagram through consumer neuroscience. *Behavioral Sciences, 11*, 108.
18. Cai, Z., Huang, Q., Liu, H., & Wang, X. (2018). Improving the agility of employees through enterprise social media: The mediating role of psychological conditions. *International Journal of Information Management, 38*, 52–63.
19. Dittes, S., Richter, S., Richter, A., & Smolnik, S. (2019). Toward the workplace of the future: How organizations can facilitate digital work. *Business Horizons, 62*, 649–661.
20. Ewing, M., Men, L. R., & O'Neil, J. (2019). Using social media to engage employees: Insights from internal communication managers. *International Journal of Strategic Communication, 13*, 110–132.
21. Goncalves, M. V., Marques, J. A. L., Silva, B. R. S., Luther, V., & Hayes, S. (2022). *Neuromarketing and global branding reaction analysis based on real-time monitoring of multiple consumer's biosignals and emotions*. Available at SSRN 4071297.
22. Gupta, V., Rubalcaba, L., Gupta, C., & Pereira, L. F. (2022). Library social networking sites for fostering startup business globalization through strategic partnerships. *The Journal of Academic Librarianship,* 102504. https://doi.org/10.1016/j.acalib.2022.102504
23. Huang, J., Baptista, J., & Newell, S. (2015). Communicational ambidexterity as a new capability to manage social media communication within organizations. *The Journal of Strategic Information Systems, 24*, 49–64.
24. Leidner, D. E., Gonzalez, E., & Koch, H. (2018). An affordance perspective of enterprise social media and organizational socialization. *The Journal of Strategic Information Systems, 27*, 117–138.
25. Men, L. R., O'Neil, J., & Ewing, M. (2020). Examining the effects of internal social media usage on employee engagement. *Public Relations Review, 46*, 101880.
26. Nivedhitha, K. S., & Sheik Manzoor, A. K. (2020). Get employees talking through enterprise social media! Reduce cyberslacking: A moderated mediation model. *Internet Research, 30*, 1167–1202.
27. Oksa, R., Saari, T., Kaakinen, M., & Oksanen, A. (2021). The motivations for and well-being implications of social media use at work among millennials and members of former generations. *International Journal of Environmental Research and Public Health, 18*, 1–22.

Media Naturalness, Emotion Contagion, and Creativity: A Laboratory Experiment Among Dyads

Maylis Saigot, Rob Gleasure, Ioanna Constantiou, and Andreas Blicher

Abstract Workplaces have evolved to rely on digital media for collaboration. Previous research has demonstrated how different characteristics of these tools, such as richness and naturalness, can enable and constrain communication among online teams. However, the role of affect in these collaborations, and the degree to which teams are able to communicate affective information, remains less clear. This research-in-progress presents a laboratory experiment that compared creative task performance under two conditions (i) dyads were online or in-person (ii) dyads began with similar or different affective states.

Keywords Digital teams · Media naturalness · Mood synchronicity · Emotion contagion

1 Introduction

Communication media allow different types of information to be communicated, including both cognitive and affective information. The *media naturalness hypothesis* suggests humans prefer face-to-face-like communication because our biological apparatus has evolved to prime us with the necessary symbolic tools and heightened physiological alertness for these face-to-face processes [1, 2]. Thus, media naturalness helps to explain some of our difficulties with digital media by showing how

M. Saigot (✉) · R. Gleasure · I. Constantiou · A. Blicher
Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark
e-mail: msa.digi@cbs.dk

R. Gleasure
e-mail: rg.digi@cbs.dk

I. Constantiou
e-mail: ic.digi@cbs.dk

A. Blicher
e-mail: abs.digi@cbs.dk

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_14

humans rely on capabilities that evolved over thousands of years to perform modern tasks [1]. Yet, despite the limitations of digital media, there is evidence that people sometimes perform as well or better using less natural media (e.g., instant messaging) rather than natural media (e.g., video conferencing).

One explanation is that collaboration is not a purely cognitive endeavor; rather, it includes a strong affective component [3]. At the offset of collaboration, team members likely exhibit different moods and affective states [4, 5]. These dissimilar moods can promote generative behavior and new possibilities as individuals confront their differences [6]. This is important, as it allows teams to harness their different perspectives and find possibilities they may not have found on their own [7, 8]. It can also lead to discomfort and emotional uncertainty among team members, who may even suffer relationship breakdowns if the affective conflict becomes engrained in their relationship [9].

Provided teams can battle through these divergent affective states, they will typically converge towards a common mood through processes of affective contagion [10, 11], which acts to reinforce specific emotions and reduce ambivalence [12–14]. The resulting shared affective "vantage point" helps team members align their expectations [15], limiting the potential distraction of exploring new possibilities.

For this reason, the ability to communicate affective information appears to be both a blessing and a curse. It is both an engine for a team's creativity and ability to push each other and a source of distraction that can create interpersonal obstacles to collaboration. We argue that, if we are to make sense of conflicting findings about teams' media preferences and performance outcomes, we must make sense of how digital media impact collaboration and allow individuals to find a suitable balance of affective conflict. Hence, this study proposes a laboratory experiment to study the effect of communication media (in-person vs. via video conference) on creativity and explore emotion contagion as a mediating factor. The remainder of this work-in-progress paper presents the research design and early results from preliminary analyses of partial data.

2 Methods

Study Design

Participants. Participants were recruited from the compensated subject pool of a West Coast American university and the study was conducted on campus. Participants were all 18 years or older, affiliated with the university, and were compensated with a $20 gift card.

Task. Participants were paired into dyads and asked to complete three consecutive trials of the so-called "alternative use task" (AUT) (e.g., [16–18]). For each trial, participants were given the name of an object and asked to generate as many alternative uses for this object as possible in 5 min. They were informed that their ideas will be evaluated based on originality, number, uniqueness, and level of detail. After instructions were displayed through a set of slides, the page automatically advanced to a blank Google Sheet with instructions and the name of the object for the ongoing trial. The order of the trials was consistent for all dyads: frisbee, newspaper, and plastic bottle. Once the 5 min were up, the page automatically advanced to the next trial.

Treatments. The experiment used a two by two, between-subjects design with four treatments: (i) in-person communication, (ii) video-conference-based communication, (iii) convergent mood, and (iv) divergent mood. Assignment to treatment was randomized. The experiment was created using Qualtrics software (www.qualtrics.com) for questionnaires, Google Sheets for task completion, and iMotions software (www.imotions.com) to program the experiment. Only one participant in each dyad was permitted to type into the Google Sheet, while the other participant could only view the shared spreadsheet. The writer's role was assigned randomly before the session and remained assigned to the same participant for all three trials.

In-person dyads completed the task face-to-face, only separated by their respective laptops, while online dyads were connected via Zoom after the mood induction. In-person dyads were brought into a shared room and seated at a round table with two workstations (14″ laptops) across from one another, separated by a divider. The divider was removed after the mood induction, right before starting the collaborative task. Online dyads were placed in separate rooms and seated at individual workstations in front of a 14″ laptop and 22″ monitor (display only) placed one behind the another. The Zoom conference was displayed on the larger monitor.

Convergent and divergent moods were manipulated via pre-task induction. Divergent moods were induced by getting the two participants of a dyad to play two different versions of a Pac-man game, one version to induce positive affect (PA-Pacman) and one to induce negative affect (NA-Pacman). Convergent moods were induced by getting both participants of a dyad to play the same version of the Pac-man game (either both positive or both negative). The game lasted five minutes, after which it stopped automatically, similar to [19]. The respective effects of the two versions of the Pac-man game were validated in a pilot of this study [20].

Instruments. Convergent and divergent moods were measured using a version of the Positive Affect and Negative Affect Scale (PANAS) administrated via a Qualtrics questionnaire before and after the mood induction. Creative performance was measured through "fluency" [21], that is, the number of ideas produced per dyad. Moreover, we used self-report measures of affective states [22], cognitive consensus [23, 24], team processes [24–28], and perceived affective friction [20]. The task outcome was measured as the mean fluency of each dyad [21].

Physiological data. During the experiment, eye gaze was measured by Tobii Pro x3-120 eye trackers, skin conductance and cardiac rhythms were recorded using Shimmer3 GSR+ and ECG, and Affectiva performs facial expression analysis (FEA).

Within each dyad, we considered a range of different analytical techniques to measure physiological synchrony as a proxy for emotional contagion, including cross-correlation [29–33], coherence, cross-recurrence, and delayed coincidence count [33, 34]. We also plan to analyze gaze overlap signals [35, 36], which have been associated with affective engagement [37, 38].

Data preparation. Prior to running statistical tests, we pre-processed the FEA data. Because each dyad performed the experimental task on two different machines, the data needed to be synchronized pairwise. iMotions provides a Unix timestamp that marks the start of data collection for each participant as well as an integrated timestamp for all sensors in milliseconds since recording started. This allowed us to derive the real-time data point for all signals and participants. However, because the data was recorded on different machines, the data was imperfectly synchronized within dyads. To allow for accurate comparison within dyads, we performed an interpolation technique on the smile coefficients and the real time of the session. We used a one-dimensional piecewise cubic Hermite interpolating polynomial [39], also known as PchipInterpolator. This technique constructs a smooth curve that passes through the given data points while maintaining monotonicity (i.e., it does not produce any local maxima or minima between data points). First, the data was filtered and cleaned, converting the 'real_time' column to numeric values. Then, the Python Scipy PchipInterpolator function [40] was used to interpolate the data for each participant. The time range was determined by finding the maximum and minimum time values for the two participants being analyzed and using this range to create a new time array with a uniform time step of 20 ms (2e7 ns). Finally, the interpolated smile values for each participant were appended to a new list that was used for the rest of the analyses.

3 Preliminary Analysis

Manipulation Check

A 2 × 2 analysis of variance (ANOVA) with time (pre-manipulation vs. post-manipulation) as a within-subjects factor and game (NA-pacman vs. PA-pacman) as a between-subject factor was conducted on the PANAS positive affect data. Post hoc dependent samples *t*-tests revealed a significant difference between post-manipulation NA-pacman dyads and pre-manipulation PA-pacman dyads, $t(47) = 3.80$, $p < 0.001$, $r = 0.12$. Overall, the results suggest that positive affect was higher after playing PA-pacman ($M = 36.0$, $SD = 8.5$) compared to pre-manipulation ($M = 34.1$, $SD = 7.4$). We also found a significant difference between post-manipulation and pre-manipulation for NA-pacman dyads, $t(51) = 3.41$, $p = 0.001$, $r = 0.17$. The results also suggest that negative affect was higher after playing NA-pacman ($M = 21.1$, $SD = 7.6$) compared to pre-manipulation ($M = 18.6$, $SD = 6.8$).

Table 1 Descriptive statistics of fluency per condition

	N	Mean	SD	SE	95% conf.	Interval
CM O	30	14.2000	3.8899	0.7102	12.7475	15.6525
CM P	24	15.7083	3.5322	0.7210	14.2168	17.1999
DM O	27	15.7778	4.1169	0.7923	14.1492	17.4064
DM P	24	16.1667	4.6966	0.9587	14.1835	18.1499

CM convergent mood, *DM* divergent mood, *O* online, *P* in person

Descriptive Statistics

We recruited 86 participants, grouped into 43 dyads (in-person: n = 23, online: n = 20, convergent mood: n = 21, divergent mood: n = 21). A total of 1998 ideas were produced across all trials, of which 145 were removed from the dataset because they either (i) did not constitute a use of the object (e.g., selling the object) or (ii) constituted a non-alternative use of the object (e.g., playing frisbee) (frisbee: n = 600, bottle: n = 668, newspaper: n = 730). Our preliminary analyses of emotional contagion used the smiling percentage of dyads based on Affectiva AFFDEX: M = 34.06 (t1, SD = 41.48), 24.36 (t2, SD = 37.53), 24.21 (t3, SD = 37.49). Table 1 provides an overview of the fluency data based on the experimental conditions.

To continue our preliminary exploration of the data, we assumed that the participant who was not writing (NW) led the smiling behavior, based on the idea that this role takes up some of the participant's attention. We performed a Wilcoxon rank-sum test to test this assumption. The results suggest that there is a significant difference in smiling scores between the writers' smiling scores and the non-writers' smiling scores ($F = -104.595$, $p < 0.001$), with non-writers showing slightly lower scores. However, the effect size of the difference was small, with a Cohen'd of −0.02. Furthermore, we visually inspected the difference in smiling scores within dyads through each trial, looking for variability patterns throughout the trials. Figure 1 suggests that participants within a dyad show greater fluctuation in the range of difference with respect to each other on a second-per-second basis. In other words, when measuring the difference between the writers' and non-writers' smiling scores each second, it seemed that this difference was more dynamic throughout Trial 1, then progressively more homogeneous as we progressed through the trials.

Preliminary Findings

We first ran a fixed effect OLS to confirm the presence of smile contagion within dyads and across trials. The results show support for emotional contagion across each of the three trials, both from the dyad member who was writing down ideas (Smile_W) on the dyad member who was not (Smile_NW), and in the opposite direction (see Table 2).

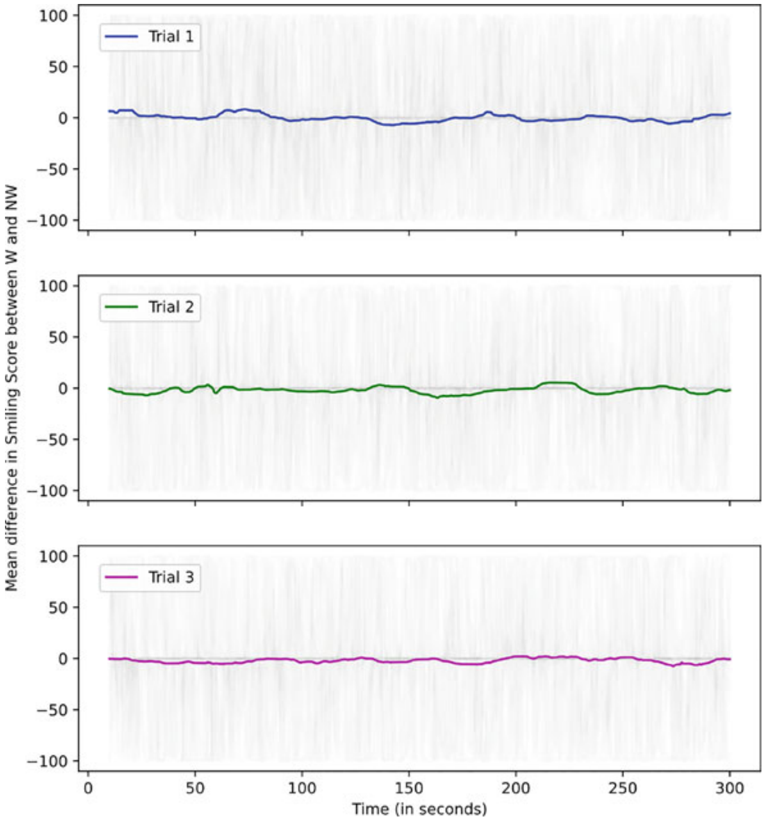


Fig. 1 Visual representation of smile variability within dyads across trials

For our preliminary analysis, we calculated a separate basic correlation score for the smiling measures for each member within a dyad. We did this by running a simple fixed effects ordinary least squares (OLS) for each dyad as follows,

$$Smile\_W_{it} = \alpha_i + \beta_1 Smile\_NW_{it} + u_{it}$$

where *Smile_W* denotes the smiling score for the team member writing down ideas, *Smiling_NW* denotes the other team member, i denotes a specific dyad, t denotes each time unit during a brainstorming session, α_i is the unobserved time-invariant individual effect, and u_{it} is the error term. We treated the smiling score for the individual who was asked to write down ideas treated as the dependent variable, and the smiling score for the other individual as the independent variable. We called this variable *Smile_contagion.*

We next performed a mixed ANCOVA using our Smile_contagion coefficient, and dummy variables derived from our experimental conditions to study their effects on fluency. The results show support for an effect of the mood condition (i.e., divergent

Table 2 Results from fixed effects OLS for contagious smiling across trials

	Dependent variable					
	Smile_NW			Smile_W		
	Frisbee	Newspaper	Plastic bottle	Frisbee	Newspaper	Plastic bottle
	(1)	(2)	(3)	(1)	(2)	(3)
Smile_W	0.357^{***} (0.009)	0.388^{***} (0.009)	0.402^{***} (0.009)			
Smily_ NW				0.345^{***} (0.009)	0.356^{***} (0.008)	0.362^{***} (0.008)
Obs.	11,060	11,064	11,114	11,060	11,064	11,114
R^2	0.123	0.138	0.146	0.123	0.138	0.146
Adj. R^2	0.121	0.135	0.143	0.121	0.135	0.143
F stat.	$1,550.929^{***}$(df = 1; 11,024)	$1,763.124^{***}$ (df = 1; 11,028)	$1,889.804^{***}$ (df = 1; 11,078)	$1,550.929^{***}$ (df = 1; 11,024)	$1,763.124^{***}$ (df = 1; 11,028)	$1,889.804^{***}$ (df = 1; 11,078)

Note $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

vs. convergent) on fluency, as well as an interaction of smile contagion and mood condition (see Table 3). This analysis suggests that dyads in the divergent mood condition are more likely to score higher on fluency, but their fluency score drops significantly when they have strong smile contagion (see Fig. 2 for visuals).

Table 3 Mixed ANCOVA results on fluency

Mixed linear model regression results						
	Coef.	Std. err.	z	P > |z|	[0.025	0.975]
Intercept	12.904	2.033	6.349	0.000	8.920	16.887
is_DM	7.776	3.321	2.342	**0.019**	−1.267	14.284
is_P	−2.450	3.253	−0.753	0.451	−8.827	3.926
Smile_contagion	4.644	6.148	0.755	0.450	−7.405	16.693
Smile_ contagion:is_DM	−19.088	9.013	−2.118	**0.034**	−36.752	−1.423
Smile_ contagion:is_P	11.744	9.040	1.299	0.194	−5.974	29.463
is_DM:is_P	−1.265	2.369	−0.534	0.593	−5.909	3.379
Group var	9.447	1.392				

Variables: *is_DM* dummy variable for divergent mood (default, the alternative is convergent mood), *is_P* dummy variable for in-person (default, the alternative is online), *Smile_contagion* coefficient

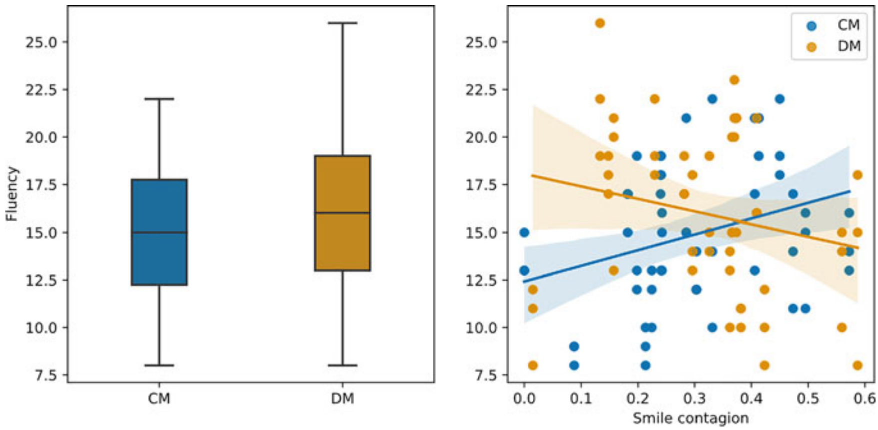


Fig. 2 Boxplot of fluency depending on the mood condition (left) and linear regression model fit for fluency versus smile contagion for the mood condition (CM = convergent mood, DM = divergent mood) (right)

Preliminary Conclusions

The early-stage results show encouraging evidence that emotional contagion could play an important role in creative fluency. This study showed that placing participants in divergent moods seemed more likely to yield greater fluency, whereas convergent dyads would produce fewer ideas. When controlling for smile contagion, however, it seemed that greater smile contagion reduced overall fluency for divergent dyads. This could be a sign that the participants have to expend more effort to overcome their emotional divergence, which may take away their attention from the task at hand—as they focus on re-establishing emotional stability within their team. These results are promising, considering that our full dataset includes other measures of creativity, including aspects such as originality, elaboration, etc. Moreover, including other variables from the questionnaires will help understand the level of awareness involved in these affective processes, as well as tangible experiences like perceived performance, quality of teamwork, etc.

4 Analysis Plan and Expected Contributions

This work-in-progress focused on reporting descriptive statistics and simple tests, allowing us to better understand the dataset and draw basic conclusions. However, this study produced a rich and complex dataset that offers great potential for further exploratory analysis. So far, the data highlighted important challenges that are unique to the interactive nature of the experiment. Among others, signal synchronization within dyads has made it difficult to calculate the co-occurrence of physiological events due to the inconsistency in physiological "leadership'—that is, there is no consistent trend as to which participant experiences physiological changes first. However, smiling data proves to be a suitable testbed for the preparation and analysis of dyadic psychophysiological data, because the signal shows low levels of noise and can easily be validated through visual inspection of the video recordings.

Plan for the Complete Analysis

We plan to continue searching for the optimal way to derive a proxy for emotional contagion in facial expression data, as well as including the rest of the sensor data that was recorded during the experiment. Specifically, we will investigate a variety of measures of emotional contagion, creativity, and quality of teamwork, taking full advantage of the richness of the data we collected thanks to our multi-modal, dyadic experimental design. Common measures of signal and physiological synchrony include cross-correlation and Mutual Information. However, these techniques both have a limited capacity to handle dynamic leadership, meaning that while they would be reliable when it is always the same participant smiling first (one leader and one follower), their derived variable become heavily biased when leadership changes dynamically throughout the experimental session (participants lead the smiling behavior interchangeably, with no consistent leader/follower).

We therefore plan to derive a variable based on smile overlap with a custom function. iMotions estimates a 50% likelihood to represent a moderately strong display of facial response. Based on this, we propose to code smiling peaks in each signal as a binary variable based on a threshold of 50 (>= 50 = 1; <50 = 0). Based on the interpolated sample for our time series, we can then filter through both signals of each dyad at an interval of 20 ms and multiply them element-wise to create a third signal that represents smile overlaps (time units where the product of the signals is equal to 1, meaning that both participants have a smiling score of at least 50 for the current time unit). Figure 3 exemplifies this process for a sample dyad randomly select in our sample. The top graph shows the signals of both participants (orange and blue). In green, we emphasize smiling peaks, which are defined as a smiling score above the threshold of 50 for both participants. The bottom graph shows the binarization of the signals and the creation of the third, overlapping signal. We can then calculate the duration of each overlapping peak by multiplying consecutive values of 1 in the third signal by our time unit of 20 ms. For each dyad, we thus obtain a count of overlapping smiling peaks (>=50) and their duration, from which we can derive the total duration of overlapping smiling behavior in seconds.

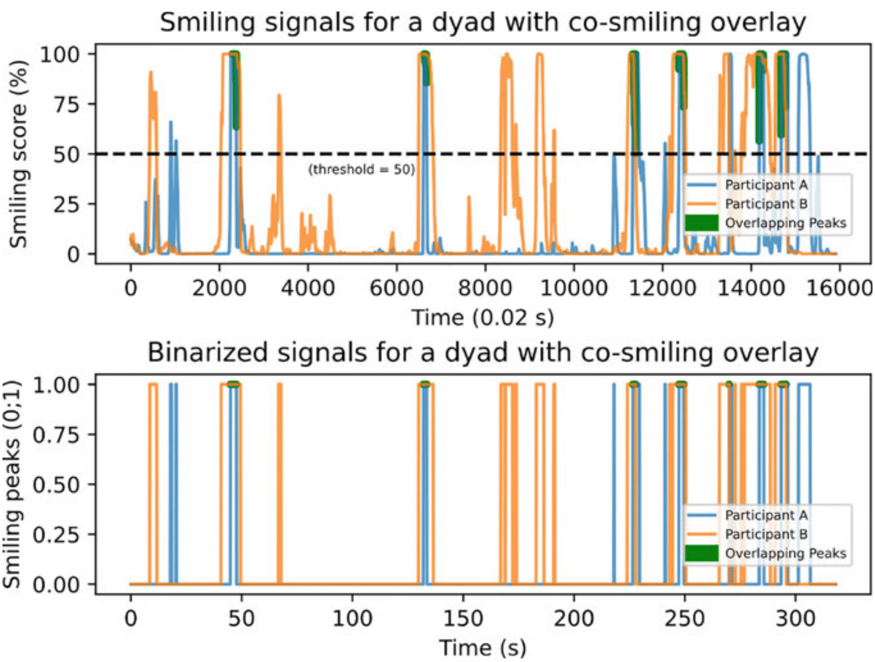


Fig. 3 Binarization of the signals and overlap detection

Although promising, this proposed variable is complex and can be sensitive to several factors. First, deriving a binary variable out of a continuous variable based on arbitrarily-defined threshold results in considerable data loss. We will thus need to carefully consider the analytical purpose of such a variable and whether 50% is a suitable threshold. Second, this technique is quite insensitive to time delays as it only accounts for synchronous peaks. This means that it could potentially overlook important information about smile contagion lag. We plan to further refine this variable to increase its accuracy, precision, and usability.

Expected Contributions

Despite the early stage of our analysis of the data, this study already shows encouraging signs of significant contributions. First, we show that placing dyads in divergent moods before the task may improve their creative fluency, as long as smile contagion is low. This finding is significant because as teams work remotely, they are more likely to be in a range of affective states, whereas when they are co-located in a shared office space, they are more likely to experience a narrower range of affective states. In other words, teams could benefit from being geographically dispersed when working on ideation tasks. If they wanted to capitalize on this advantage, they might achieve even better results when prioritizing communication media that make it harder for emotions to spread. Such media could be those types that were traditionally considered lean in the media richness and synchronicity literature—although more research is needed to specifically investigate the affective nature of communication media.

Second, we propose an experimental design to study dyadic interaction in a laboratory while using a naturalistic protocol. Using applications like Zoom and Google Sheets is uncommon in laboratory experiments. While these tools presented some limitations in terms of experimental control, we chose to prioritize ecological validity by selecting tools that are already commonly used by teams in organizations. Our research design and protocol contribute to expanding the applicability of laboratory experiments in the fields of business and management.

Third, we suggest new directions for preparing and analyzing dyadic psychophysiological data. After we complete our analysis, we plan to make our data processing pipeline publicly available. In doing so, we want to encourage other scholars to pursue dyadic psychophysiological experiments involving physiological synchrony.

References

1. Kock, N. (2004). The psychobiological model: Towards a new theory of computer-mediated communication based on Darwinian evolution. *Organization Science, 15*, 327–348. https://doi.org/10.1287/orsc.1040.0071

2. Kock, N. (2011). Media naturalness theory: Human evolution and behaviour towards electronic communication technologies. In *Applied evolutionary psychology*. OUP Oxford.
3. Amabile, T., Barsade, S., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly, 50*, 367–403. https://doi.org/10.2189/asqu.2005.50.3.367
4. Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In *Handbook of emotions* (pp. 381–403). The Guilford Press.
5. Weiss, H. M., & Cropanzano, R. (1996). Affective events theory: A theoretical discussion of the structure, cause and consequences of affective experiences at work. *Research in organizational behavior, 18*, 75.
6. Jarvenpaa, S. L., Standaert, W., & Vlerick Business School. (2018). Digital probes as opening possibilities of generativity. *JAIS,* 982–1000. https://doi.org/10.17705/1jais.00516
7. Barsade, S., & Knight, A. P. (2015). Group affect. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 21–46. https://doi.org/10.1146/annurev-orgpsych-032414-111316
8. Fong, C. T. (2006). The effects of emotional ambivalence on creativity. *Academy of Management Journal, 49*, 1016–1030. https://doi.org/10.5465/AMJ.2006.22798182
9. Barsade, S., Ward, A. J., Turner, J. D. F., & Sonnenfeld, J. A. (2000). To your heart's content: A model of affective diversity in top management teams. *Administrative Science Quarterly, 45*, 802–836. https://doi.org/10.2307/2667020
10. Barsade, S., Coutifaris, C. G. V., & Pillemer, J. (2018). Emotional contagion in organizational life. *Research in Organizational Behavior, 38*, 137–151. https://doi.org/10.1016/j.riob.2018.11.005
11. Hatfield, E., Cacioppo, J., & Rapson, R. (1992). Primitive emotional contagion. In *Current directions in psychological science* (pp. 151–177).
12. Bettenhausen, K. L. (1991). Five years of groups research: What we have learned and what needs to be addressed. *Journal of Management, 17*, 345–381. https://doi.org/10.1177/014920639101700205
13. Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly, 44*, 741–763. https://doi.org/10.2307/2667054
14. Zhu, D. H. (2013). Group polarization on corporate boards: Theory and evidence on board decisions about acquisition premiums. *Strategic Management Journal, 34*, 800–822. https://doi.org/10.1002/smj.2039
15. Vuori, T. O., & Huy, Q. N. (2016). Distributed attention and shared emotions in the innovation process: How Nokia lost the smartphone battle. *Administrative Science Quarterly, 61*, 9–51. https://doi.org/10.1177/0001839215606951
16. Brucks, M. S., & Levav, J. (2022). Virtual communication curbs creative idea generation. *Nature*. https://doi.org/10.1038/s41586-022-04643-y
17. Harada, T. (2020). The effects of risk-taking, exploitation, and exploration on creativity. *PLoS ONE, 15*, e0235698. https://doi.org/10.1371/journal.pone.0235698
18. Klein, A., & Badia, T. (2022). Where divergent ideas converge: Answers to AUT found on short list of word co-occurrences terms. *Creativity Research Journal,* 1–17. https://doi.org/10.1080/10400419.2022.2103314
19. Reuderink, B., Nijholt, A., & Poel, M. (2009). *Affective Pacman: A frustrating game for brain-computer interface experiments* (pp. 221–227).
20. Saigot, M. (2022). Leveraging affective friction to improve online creative collaboration: An experimental design. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 237–250). Springer International Publishing.
21. Guilford, J. P. (1975). Varieties of creative giftedness, their measurement and development. *Gifted Child Quarterly, 19*, 107–121. https://doi.org/10.1177/001698627501900216
22. Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

23. Mohammed, S., & Ringseis, E. (2001). Cognitive diversity and consensus in group decision making: The role of inputs, processes, and outcomes. *Organizational Behavior and Human Decision Processes, 85*, 310–335. https://doi.org/10.1006/obhd.2000.2943
24. O'Leary, K., Gleasure, R., O'Reilly, P., & Feller, J. (2022). Introducing the concept of creative ancestry as a means of increasing perceived fairness and satisfaction in online collaboration: An experimental study. *Technovation, 110*, 102369. https://doi.org/10.1016/j.technovation.2021.102369
25. Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. *Organization Science, 12*, 435–449. https://doi.org/10.1287/orsc.12.4.435.10635
26. Nunamaker, J. F., Applegate, L. M., & Konsynski, B. R. (1987). Facilitating group creativity: Experience with a group decision support system. *Journal of Management Information Systems, 3*, 5–19. https://doi.org/10.1080/07421222.1987.11517775
27. Valacich, J. S., Dennis, A. R., & Nunamaker, J. F. (1992). Group size and anonymity effects on computer-mediated idea generation. *Small Group Research, 23*, 49–73. https://doi.org/10.1177/1046496492231004
28. Wu, I.-L., & Chiu, M.-L. (2018). Examining supply chain collaboration with determinants and performance impact: Social capital, justice, and technology use perspectives. *International Journal of Information Management, 39*, 5–19. https://doi.org/10.1016/j.ijinfomgt.2017.11.004
29. Chanel, G., Kivikangas, J. M., & Ravaja, N. (2012). Physiological compliance for social gaming analysis: Cooperative versus competitive play☆. *Interacting with Computers, 24*, 306–316. https://doi.org/10.1016/j.intcom.2012.04.012
30. Dehais, F., Vergotte, G., Drougard, N., Ferraro, G., Somon, B., Ponzoni Carvalho Chanel, C., & Roy, R. (2021). *AI can fool us humans, but not at the psycho-physiological level: A hyperscanning and physiological synchrony study.*
31. Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. *Applied Ergonomics, 40*, 997–1003. https://doi.org/10.1016/j.apergo.2009.02.002
32. Järvelä, S., Kivikangas, J. M., Kätsyri, J., & Ravaja, N. (2014). Physiological linkage of dyadic gaming experience. *Simulation & Gaming, 45*, 24–40. https://doi.org/10.1177/1046878113513080
33. Verdiere, K. J., Albert, M., Dehais, F., & Roy, R. N. (2020). Physiological synchrony revealed by delayed coincidence count: Application to a cooperative complex environment. *IEEE Transactions on Human-Machine Systems, 50*, 395–404. https://doi.org/10.1109/THMS.2020.2986417
34. Mønster, D., Håkonsson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & Behavior, 156*, 24–34. https://doi.org/10.1016/j.physbeh.2016.01.004
35. Schneider, B., & Pea, R. (2017). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. In M. Orey & R. M. Branch (Eds.), *Educational media and technology yearbook* (pp. 99–125). Springer International Publishing.
36. D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6245–6290). ACM.
37. Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00711
38. Janssen, J. H., Bailenson, J. N., IJsselsteijn, W. A., & Westerink, J. H. D. M. (2010). Intimate heartbeats: Opportunities for affective communication technology. *IEEE Transactions on Affective Computing, 1,* 72–80. https://doi.org/10.1109/T-AFFC.2010.13
39. Fritsch, F. N., & Butland, J. (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific Computing, 5,* 300–304. https://doi.org/10.1137/0905021

40. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17,* 261–272. https://doi.org/10.1038/s41592-019-0686-2

The Role of User Control in Enhancing Human-AI Collaboration Effectiveness: Insights from a Pilot Study

Burak Oz, Alexander Karran, Jared Boasen, Constantinos Coursaris, and Pierre-Majorique Léger

Abstract In this research program proposal, we aim to investigate why experts override AI suggestions and identify design principles for more effective human-AI teams. Specifically, we propose testing whether increasing the perceived locus of control of human decision-makers over AI functions will lead to fewer overrides and improved performance. We present a mixed-factorial, multi-trial experimental design in which participants receive AI recommendations regarding demand forecasting decisions in a business simulation. Prior to each trial, one group specifies how they want the AI to function (experimental), and the other group does not (control). We use electroencephalography and oculometry to capture attention to recommendations and user interface elements. Behavioral data from a preliminary pilot study with four participants align with our hypotheses. We observed that participants in the experimental condition applied smaller adjustments to AI suggestions and had higher decision performance than the control group. The experiment's results will contribute to our understanding of AI aversion and inform the design of human-AI interactions to improve performance.

B. Oz (✉) · A. Karran · J. Boasen · C. Coursaris · P.-M. Léger
Tech3Lab, HEC Montréal, Montréal, Québec, Canada
e-mail: burak.oz@hec.ca

A. Karran
e-mail: alexander.karran@hec.ca

J. Boasen
e-mail: jared.boasen@hec.ca

C. Coursaris
e-mail: constantinos.coursaris@hec.ca

P.-M. Léger
e-mail: pml@hec.ca

J. Boasen
Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_15

Keywords AI aversion · Artificial intelligence · Decision-making · Attention · Perceived locus of control · Human-AI interaction · NeuroIS · EEG · ERP · P300 · Eye-tracking

1 Introduction

Data-driven intelligent systems empowered by artificial intelligence (AI)[1] methods are increasingly being implemented in various fields [1–4]. However, the implementation potential of AI in complex decision-making tasks is far from full automation, primarily due to challenges in applying algorithmic solutions in critical contexts where there are concerns regarding ethical, social, or life-critical aspects of the final decision [5–9]. These concerns regarding full automation have given rise to semi-autonomous human-AI teams, where labor is divided between humans and AI. These systems involve collaboration between human experts and AI systems to perform tasks that require both human intuition and judgment abilities for dealing with unstructured problems and machine capabilities of big data analytics for dealing with structured problems [10, 11]. However, reports of AI implementation in knowledge work have shown that expert decision-makers often override AI suggestions and end up with worse decision performance [12–15].

Experts' reluctance to accept AI suggestions, even when AI performs well, can be referred to as AI aversion, akin to algorithm aversion [13]. AI aversion has been attributed to numerous factors such as lack of transparency, fear of bias, lack of accountability, preference for human judgment, and influence of personal experiences leading to different user expectations [16]. In essence, AI aversion occurs because of a lack of cognitive compatibility between human experts and intelligent systems, especially when these systems lack transparency and configurability [16–18].

Increasing compatibility between humans and AI systems can be possible through two primary methods: increasing human control over the system and improving system explanations [16, 19]. Research on AI explainability is vast and has produced valuable insights. On the other hand, although there is growing interest in studying user control, there is still a need for a better understanding of the impact of varying types and degrees of users' perceived locus of control on the effectiveness and acceptance of AI systems [19]. In this study, we define perceived locus of control as a person's perception of their ability to influence the human-AI semi-autonomous decision-making process [19, 20]. As part of a new research program investigating experts' AI aversion, we propose research investigating the impact of perceived locus of control on users' AI acceptance. More specifically, the proposed study aims to address the following research question:

[1] We employ the term "AI" to encompass a broad range of automation and autonomy capabilities, irrespective of the underlying technology, which could include learning agents like machine learning, deep learning, or artificial intelligence, as well as more static reasoning systems.

RQ1: To what extent does a user's perceived locus of control improve human-AI team performance?

While reducing AI aversion is critical, there is an equally important and relevant concept on the other end of the spectrum: algorithm overreliance [21]. Some experts may rely too heavily on algorithmic suggestions without considering contextual nuances and outliers, defying the purpose of implementing a human-AI team. Indeed, researchers express concerns that overreliance on AI can decrease system performance and error detection [22]. However, these are yet to be empirically tested:

RQ2: To what extent does overreliance on AI suggestions impact a user's ability to detect AI errors?

To address recent calls for research to leverage NeuroIS methods to develop more effective human-AI teams [22] and generate prescriptive IS design knowledge [23], we started a pilot study involving a mixed factorial, multi-trial experiment consisting of a demand forecasting task. We use electroencephalography (EEG) and oculometry to measure attention, cognitive processing, and gaze behaviors. We manipulate AI configurability across participants for the first research question and AI performance across trials for the second research question. We hypothesize that subjects permitted to specify AI function parameters will have a higher perceived locus of control, override the AI less often, exhibit stronger attentional processing and gaze behaviors, and perform the task better compared to the control group, even if the AI system provides the same suggestions regardless of user input. The findings of this study are expected to provide a better understanding of human-AI interactions that can be useful in informing developers on the design of effective and efficient human-AI teams.

In this submission, we report on our study design, briefly discuss preliminary results from a pilot study comprised of four participants and conclude with the expected contributions of the complete study currently underway.

2 Methods and Materials

To address the aims of this research plan, we developed an experimental task in which participants are asked to perform a series of demand forecasting tasks. This section explains the experimental task, followed by the experimental conditions, procedure, and data recording tools.

Participants

We will use convenience sampling to recruit 100 participants, primarily from our university's student participant panel. Participants will be screened based on their experience with SAP and the courses they have taken to ensure a basic knowledge of the experimental task and experience with stimuli. In addition to their task-related knowledge, participants will be screened for having a normal vision and not being diagnosed with a neuropathological condition.

Material and Procedure

Building upon previous studies using the ERPsim business simulation to investigate human–computer interactions [24–28], we used this platform to generate a realistic demand dataset for the demand forecasting task. The experimental task uses the newsvendor problem, a well-established experimental paradigm in which participants play the role of a demand management agent [29, 30]. They decide how many units of a perishable product to order for the next period, knowing there is uncertain demand and different costs for having excess demand or losing sales.

Participants will be presented with a time series chart displaying a fictitious distribution company's weekly ice cream demand over the last 20 weeks. In each trial, they will evaluate the weekly demand data, receive a forecast estimation from the AI, and decide whether they want to edit the AI suggestion or continue directly. After they submit their decision, the simulation will advance by one week, and they will be presented with their performance in terms of lost and saved money. Following the performance report, participants continue doing this demand forecasting task in the same market, one virtual week per trial. With the addition of AI functionality to the newsvendor paradigm, we extend and demonstrate the use of this well-established decision-making task in an information systems study.

To ensure the ecological validity of the task, participants will use an industry-standard organizational resource planning software, SAP, to complete their experimental tasks. The SAP screen designed for this experiment shows pricing information, past demand data, and AI-supported demand forecasting functions. At the beginning of a trial, participants will not be provided with any AI suggestions or explanations, but the AI suggestion field will be visible without any number. Participants will be asked to scan and evaluate all demand information first, then fixate on the suggestion area while clicking the "Calculate AI Suggestions" button. One second after they click the button, a three or four-digit number will appear as an AI suggestion in its designated area, marking the onset of each trial's stimulus. Two seconds later, an explanation of the factors contributing to each suggestion will appear, together with buttons to accept or edit suggestions.

The designed SAP interface will be shown in full screen mode on a 22″ screen at 1920×1080 resolution at a 60 Hz refresh rate. Participants will mainly use a mouse to interact with the system and use the keyboard only when they adjust AI suggestions.

Experiment design. To test the impact of perceived locus of control on user behaviors in an AI-supported decision-making scenario, we designed a mixed factorial, multi-trial experiment with a between-subjects, two-level factor of AI configurability, and a within-subjects two-level factor of AI error.

Participants receive AI suggestions in all between-subjects conditions before completing their tasks. To investigate the impact of perceived user control on acceptance of AI suggestions and human-AI interactions, we aim to create an increased locus of control for participants in one of the between-subjects conditions. In the configurable AI (experimental) condition, the system asks for the user's configuration input at the beginning of each trial. In contrast, the unconfigurable AI (control) condition uses an AI system that does not ask for user input. Regardless of user inputs, both systems perform identically to isolate the psychological impact of the locus of control over performance. Participants in both groups can override AI suggestions.

The three AI parameters we ask participants to configure are smoothing, trend, and seasonality. These parameters are crucial in the Holt-Winters method, a popular time series forecasting model used in supply chain management [31]. We hypothesize that ability to provide input regarding relevant task parameters will increase their locus of control, which then will lead to a greater acceptance and utilization of AI suggestions in trials with high AI performance. Moreover, by making participants think about these parameters in each trial, we aim to increase their AI error detection abilities through increased task engagement and accumulated task knowledge.

We will conduct manipulation checks using previously validated self-report measurement item sets for perceived autonomy in human-AI interactions [19]. Moreover, to control the experiment duration across conditions, participants assigned to the control group will see a loading page for a duration that is based on our pretest observations.

In both between-subjects conditions, we follow an oddball paradigm design and randomly assign 15 of the 70 trials as oddballs in which the AI misbehaves, i.e., provides suggestions having more than three times the standard error of its average performance.

Experimental procedure. The experiment starts with a practice round in which participants are asked to complete five trials of demand forecasting tasks without AI support. After the practice block, there will be a training block with AI functionality, which is then followed by 70 trials of AI-supported demand forecasting tasks.

Data Recording and Analysis

Neurophysiological measures of attention to AI recommendations and the interface. EEG signals will be recorded from 32 scalp sites at a sampling rate of 1000 Hz. From this recorded data, event-related potentials (ERP) will be derived from the visual onset of the AI recommendation (the three to four-digit number at the bottom of Fig. 1). In this study, we will focus on the P300 ERP component, which reflects the cognitive processing of decision-relevant information and the allocation of attention to the stimulus [32].

The recorded EEG data will be filtered using 30 Hz low-pass and 1 Hz high-pass filters. The filtered data will then be divided into segments of 1 s per trial, starting 0.1 s

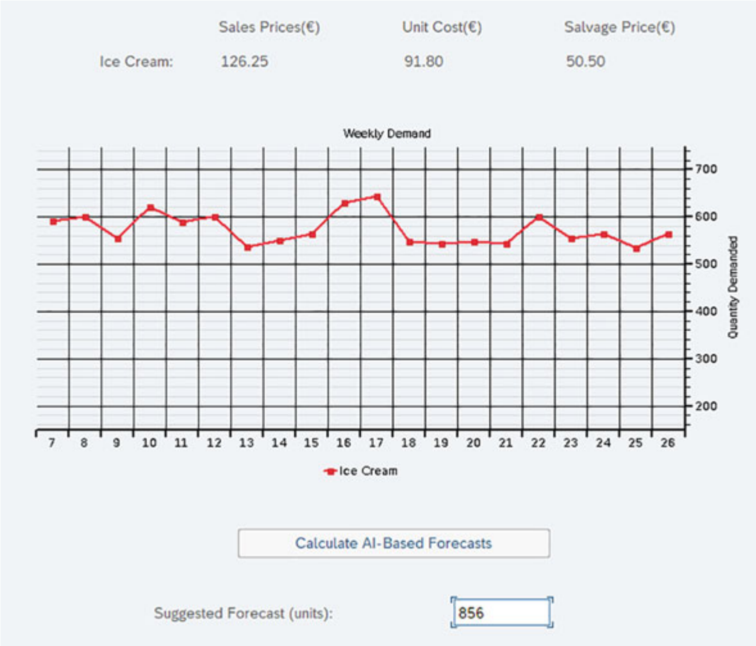


Fig. 1 Experimental stimuli. The relevant input information for the decision is on the top, whereas the AI-supported decision-making interface elements are on the bottom half of the screen. Participants are asked to analyze the data before clicking the "Calculate AI-Based Forecasts" button. They are also asked to focus on the suggestion field at the bottom. Once they receive suggestions, participants can accept or edit the suggested forecasts

before and 0.9 s after the stimulus onset of AI suggestion numbers. Following the epoching of the data, artifact rejection will be applied using automatic and manual procedures, assuming peak-to-peak amplitude ranges greater than 100 μV to be artifacts.

After filtering, epoching, and artifact rejection, the ERPs will be averaged across trials. These averaged ERPs will be used in statistical tests to determine whether our manipulation leads to a higher P300 peak amplitude, indicating higher attention to AI suggestions and improved ability to detect AI errors.

Eye-tracking measures. Gaze data will be recorded using a Tobii Pro Nano (Tobii Technology AB, Sweden) at a sampling rate of 60 Hz. Research has demonstrated that gaze transition entropy (GTE), a measure of the randomness of eye movements during visual processing, is related to attentional processing [33, 34]. Using GTE, we aim to gain insight into how individuals allocate attention when making forecasting decisions and detecting errors, and how this process is impacted by their perceived control over the AI system. Additionally, eye-tracking data will complement the ERP measurements by verifying participants' fixation on the stimulus at the onset.

Behavioral measures. We will also collect behavioral and self-report measures to gain a more comprehensive understanding of the phenomenon and triangulate our neurophysiological measurements. Behavioral measures include forecasting accuracy, response times, AI suggestion acceptance, and the magnitude of adjustments to AI suggestions. Self-report measures include the perceived level of control [19, 35, 36], perceived use and ease of use [37], trust in AI [38, 39], explanation satisfaction and understandability [39–41] and confirmation of expectations [40].

3 Preliminary Results and Future Outlook

We have conducted four pilot study sessions and analyzed behavioral and self-reported data. These results provide promising initial figures. First, our measurement of perceived level of control [19] supports the effectiveness of our manipulation. Second, our pilot study participants with a higher sense of control (i) made smaller adjustments to the AI suggestions, and (ii) obtained a higher forecasting performance. More specifically, in a time series with a seasonality ranging from 400 to 1,000 units, participants in the experimental group had an average adjustment of 128 units, whereas the control group participants had an AI adjustment average of 149 units across 70 trials. Also, participants in the experimental condition generated an average profit of 699 Euros (based on the game scenario), compared to 396 Euros by the control (140 observations per condition). Analysis of the ERP data from these pilot sessions is currently in progress.

This study is expected to provide empirical evidence for the relevance and importance of users' perceived sense of control over an AI in the success of human-AI collaborative system implementation in the context of decision-making. Moreover, the insights gained from the results of this study will have practical implications for the design of human-AI interactions. We believe that utilizing neurophysiological measures will enhance our understanding of the cognitive processes underlying human-AI interactions, leading to the design of more effective systems that optimize the contributions of both AI and human experts in their collaborative efforts.

References

1. Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*, 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005
2. Currie, G., Hawk, K. E., Rohren, E., Vial, A., & Klein, R. (2019). Machine learning and deep learning in medical imaging: Intelligent imaging. *Journal of Medical Imaging and Radiation Sciences, 50*, 477–487.
3. Fourcade, A., & Khonsari, R. (2019). Deep learning in medical image analysis: A third eye for doctors. *Journal of Stomatology, Oral and Maxillofacial Surgery, 120*, 279–288.

4. Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology, 15*, 512–520.
5. Anderson, P. L. (2019). *Damages caused by AI errors and omissions: Management complicity, malware, and misuse of data.* Anderson Economic Group.
6. Cheatham, B., Javanmardian, K., & Samandari, H. (2019). Confronting the risks of artificial intelligence. *McKinsey Quarterly, 2*, 38.
7. Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access, 6*, 14410–14430.
8. Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine, 40*, 72–80.
9. Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., & Kitano, H. (2022). *Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report*. arXiv:2210.15767
10. Endsley, M. R. (2022). Supporting human-AI teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior,* 107574. https://doi.org/10.1016/j.chb.2022.107574
11. Rai, A., Constantinides, P., & Sarker, S. (2019). Next generation digital platforms: Toward human-AI hybrids. *MIS Quarterly, 43*, iii–ix.
12. Allen, R., & Choudhury, P. (Raj). (2022). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science, 33*, 149–169. https://doi.org/10.1287/orsc.2021.1554
13. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*, 114–126. https://doi.org/10.1037/xge0000033
14. Khosrowabadi, N., Hoberg, K., & Imdahl, C. (2022). Evaluating human behaviour in response to AI recommendations for judgemental forecasting. *European Journal of Operational Research, 303*, 1151–1167. https://doi.org/10.1016/j.ejor.2022.03.017
15. Kesavan, S., & Kushwaha, T. (2020). Field experiment on the profit implications of merchants' discretionary power to override data-driven decision-making tools. *Management Science, 66*, 5182–5190. https://doi.org/10.1287/mnsc.2020.3743
16. Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*, 220–239. https://doi.org/10.1002/bdm.2155
17. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
18. European Commission. (2021). *Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.*
19. Westphal, M., Vössing, M., Satzger, G., Yom-Tov, G. B., & Rafaeli, A. (2023). Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior,* 107714. https://doi.org/10.1016/j.chb.2023.107714
20. Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*, 1.
21. Grissinger, M. (2019). Understanding human over-reliance on technology. *Pharmacy and Therapeutics, 44*, 320.
22. Rubin, D. L. (2019). Artificial intelligence in imaging: The radiologist's role. *Journal of the American College of Radiology, 16*, 1309–1317. https://doi.org/10.1016/j.jacr.2019.05.036
23. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29,* 9–24. https://doi.org/10.1080/0960085X.2019.1708218

24. Léger, P.-M. (2006). Using a simulation game approach to teach enterprise resource planning concepts. *Journal of Information Systems Education, 17*, 441–447.
25. Léger, P., Robert, J., Babin, G., Pellerin, R., & Wagner, B. (2007). *ERPsim.* ERPsim Lab (erpsim.hec.ca), HEC Montreal.
26. Oz, B., Tran-Nguyen, K., Coursaris, C. K., Robert, J., & Léger, P.-M. (2020). Using digital nudges on analytics dashboards to reduce anchoring bias. In *SIGHCI 2020 Proceedings* (p. 3).
27. Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior, 34,* 273–283. https://doi.org/10.1016/j.chb.2014.02.011
28. Karran, A., Demazure, T., Léger, P.-M., Labonte-LeMoyne, E., Sénécal, S., Fredette, M., & Babin, G. (2019). Towards a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. *Frontiers in Human Neuroscience, 13*. https://doi.org/10.3389/fnhum.2019.00393
29. Whitin, T. M. (1955). Inventory control and price theory. *Management Science, 2*, 61–68. https://doi.org/10.1287/mnsc.2.1.61
30. Benzion, U., Cohen, Y., Peled, R., & Shavit, T. (2008). Decision-making and the newsvendor problem: An experimental study. *Journal of the Operational Research Society, 59*, 1281–1287. https://doi.org/10.1057/palgrave.jors.2602470
31. Supriatna, A., Hertini, E., Saputra, J., Subartini, B., & Robbani, A. A. (2019). The forecasting of foreign tourists arrival in Indonesia based on the supply chain management: An application of artificial neural network and Holt Winters approaches. *International Journal of Supply Chain Management, 8*, 156.
32. Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems, 37*. https://doi.org/10/ghtd97
33. Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*, 3–55.
34. Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews, 96*, 353–366.
35. Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly: Management Information Systems, 24*, 665–694. https://doi.org/10.2307/3250951
36. Tapal, A., Oren, E., Dar, R., & Eitam, B. (2017). The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in Psychology, 8*, 1552. https://doi.org/10.3389/fpsyg.2017.01552
37. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319–340. https://doi.org/10.2307/249008
38. Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*, 53–71.
39. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science, 5*.
40. Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly, 25*, 351–370. https://doi.org/10.2307/3250921
41. Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society, 23*, 128–147.

Expanding the Scope — Cognitive Robotics Meets NeuroIS

Renan Lima Baima, Letícia Mara Berto, and Tamara Roth

Abstract Cognitive Robotics aims to develop robots that can perform tasks, learn from experiences, and adapt to new situations using cognitive skills. Rooted in neuroscience theories, Cognitive Robotics provides a unique opportunity for NeuroIS researchers to theorize and imagine intelligent autonomous agents as natural cognitive systems. By translating Cognitive Robotics methods and architectures into the NeuroIS into the 2×2 design science research matrix, we intend to help researchers gain deeper insights into how humans perceive and interact with their environment. These insights may not only improve cognitive architectures but may also enable a better design and evaluation of user-centric NeuroIS systems, safer test propositions, and better self-adaptable systems that can effectively collaborate with humans in various settings.

Keywords Cognitive robotics · NeuroIS · Cross-fertilization · Design science research · Cognitive architecture

R. L. Baima (✉) · T. Roth
FINATRAX, Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Esch-sur-Alzette, Luxembourg
e-mail: renan.limabaima@uni.lu; renan.baima@ic.unicamp.br

T. Roth
e-mail: TRoth@walton.uark.edu

Present Address:
R. L. Baima · L. M. Berto
LaRoCS, Institute of Computing, University of Campinas (UNICAMP), São Paulo, Brazil
e-mail: leticia.berto@ic.unicamp.br

Present Address:
T. Roth
Sam M. Walton College of Business, University of Arkansas, Fayetteville, USA

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_16

1 Introduction

NeuroIS has proven its value as a bridge between neuroscience, psychology, and information systems research to study the impact of new technology and its use. The data obtained from the research process can also inform new designs and applications of information systems [1]. NeuroIS researchers typically use self-reporting data to explore the effects of technology use in addition to neuroimaging techniques such as EEG, fMRI, and eye tracking to collect data on brain activity while participants interact with technology [2]. This helps them gain richer insights into how users perceive and process information, make decisions, and experience emotions when using technology. Results of these analyses often deliver objectives or requirements that can be used to improve or design information systems focused on the user [3, 4].

Since NeuroIS research heavily draws on neurophysiological data to test hypotheses, a substantial number of participants is required for each experiment [1, 5]. However, it is often difficult to find the required number of participants that meet sampling criteria [6]. At the same time, neurophysiological methods require a joint analysis of environmental stimuli, the neural system, and bodily reactions, which is why virtual models cannot easily replace participants. Recent advances in cognitive neuroscience, however, may push this boundary. Gravish and Lauder [7], for instance, elaborated on using robots as surrogates to study behavior and cognition in a controlled environment. Considering robots as a model of the living system under investigation results in exploring an often isolated albeit very concrete (behavioral) variable [7]. Testing such variables in vivo and in a natural environment is challenging. Thus, using robots as a model of the living system under investigation enables researchers to make inferences about the living system that would otherwise have been difficult to obtain [8]. At the same time, robot-supported social cognitive neuroscience (rSCN) uses robots as a new type of stimuli to study cognition and behavior in humans and animals in a highly controlled experimental setting [9].

Although these more technical advancements in neuroscience already provide promising avenues of research for NeuroIS, another emerging field in robotics may push the boundaries even further. Cognitive Robotics (CogRob) uses innovation in robotics, artificial intelligence, and cognitive science to design robots that can perform complex tasks autonomously and adapt to changing environments [10]. These robots can be used in various settings, such as manufacturing, healthcare care [11], and space exploration [12]. They often help improve efficiency and reduce costs [13].

CogRob typically combines neuroscience and engineering with other disciplines, such as psychology and social sciences. This elevates robots beyond their use as simple tools or models. Robots can become platforms that help researchers explore complex cognitive issues in human-technology interaction in various social contexts [14]. Moreover, insights from CogRob can benefit the development of more powerful learning algorithms that enable the study of controlled variables in isolation providing new angles for research in psychology and neuroscience. CogRob also allows for the development of more robust robots to effectively collaborate with humans [8, 14, 15]. To better leverage the advancements in CogRob for NeuroIS, we propose CogRob methods as integral elements of the NeuroIS Design Science Cycle (DSC). We specifically aim to investigate how integrating CogRob methods and models into the NeuroIS DSC can help develop more user-centric information systems. In the following sections, we will explore the background of CogRob, its foundation, the respective methods, and how CogRob can help create better experimental settings for NeuroIS research and aid the design of user-centric information systems within the DSC.

2 Cognitive Robotics as an Emerging Field

Cognitive Robotics is an interdisciplinary field that aims to develop robots that can perform tasks, adapt to new situations [16], and learn from experiences to create machines inspired by how humans think and learn [10]. CogRob uses cognitive skills such as memory, decision-making, action understanding, and prediction but also combines ideas from other fields, such as computer science, robotics, artificial intelligence, psychology, neuroscience, and philosophy. The intention of CogRob is not to replace but to efficiently learn from problem-solving human interactions [17].

Relevant architectures and models at the intersection of neuroscience and robotics were built on the initial memory theory that surrounds, for example, the Simon and Feigenbaum architecture for cognition. The EPAM model [18] includes learnings from human memory and speech development. Later Anderson, who researched human memory, proposed the Human Associative Memory (HAM) model [19], which his student Bower further developed into the ACT model [20]. ACT-R is a cognitive architecture that aims to explain how humans perform tasks, learn new skills, and solve problems, including aspects of long-term memory and thinking processes.

Researchers use architectures and models to develop robots that can learn and reason like humans. This may help them gain insights into how humans process information, interact with their environment, and adapt to new situations [21, 22]. It can also improve the general understanding of human cognition and behavior, which may provide the foundation for the development of new therapies and interventions for persons with cognitive impairments or disabilities neuroscience [22],

The theoretical foundations of CogRob are based on learning theories such as reinforcement learning, unsupervised learning, and imitation learning [23]. Many methods and approaches used in CogRob are known from artificial intelligence machine learning and natural language processing research. The main models and architectures in CogRob include behavior-based robotics, hybrid architectures, and cognitive architectures [24], which are commonly symbolic, connectionist, or hybrid [25]. The architectures typically follow a bottom-up approach in which basic rules or nodes generate complex behavior. This differs from common AI approaches, in which a top-down process created by the programmer inspires behavior [26]. More specifically, CogRob builds on the assumption that the mind has various modular cognitive units, each responsible for a specific aspect of human cognition [20]. These interconnected modules work together to generate intelligent behavior [27]. Anderson's work [20], "The Architecture of Cognition," is considered a landmark in this context. Kotseruba and Tsotsos [24] build on his work and say that combining psychology and computer science insights inspired the first cognitive architectures, while theoretical models of human cognitive processes and related software artifacts inspired CogRob.

3 Generalizing the Cognitive Robotics Approach

In Cognitive Robotics, robots typically undergo a cognitive cycle similar to that of humans. This cycle starts with sensing the environment through sensors, interpreting the data received during the perceptual process, making decisions based on internal needs, possible actions, outcomes, emotions, motivations, and context, and acting in the environment. Figure 1 illustrates a basic cognitive cycle with possible submodules for each component of cognition.

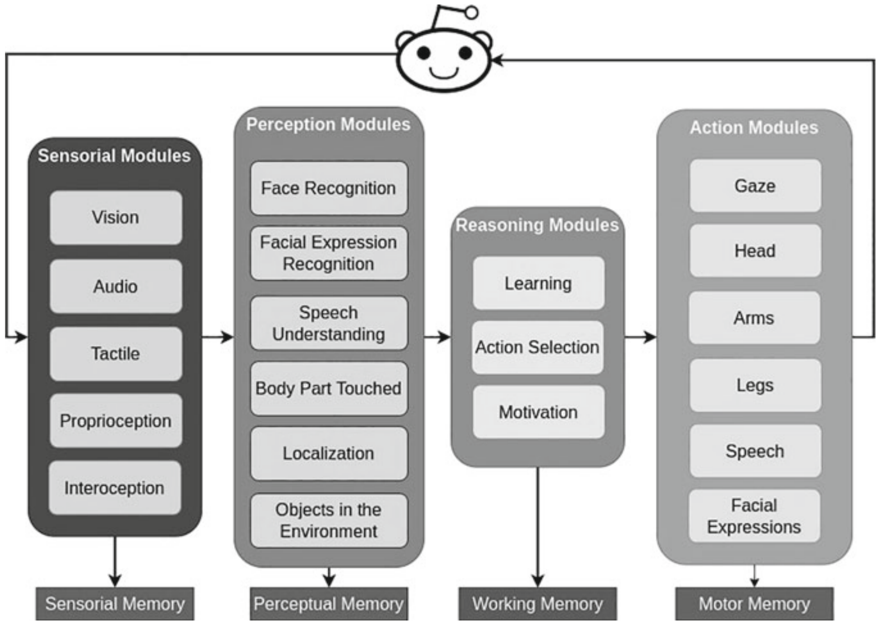


Fig. 1 Basic cognitive cycle used in a cognitive architecture. The submodules in each module are examples of possible components the robot can use. Each developer can choose which one to use and how to implement it

All modules can run in parallel according to their specificity. This allows each module to be modeled individually and then connected to the system, which is particularly valuable for exploring the impact of one module in the application without changing the entire cognitive architecture. Isolating the modules makes it possible to work in one specific module and observe the agent's behavior with certain changes. This enables accurate comparisons between approaches while maintaining the cognitive process. The work in Berto et al. [28], and strategies a simulated autonomous agent learned using two motivational mechanisms as part of the cognitive architecture. In this way, researchers could investigate and compare the impact of different elements, such as needs and pleasure, during the decision-making process under the same conditions.

4 Cognitive Robotics and the NeuroIS Design Science Cycle

Potential cross-fertilization between neuroscience and Cognitive Robotics particularly focused on how CogRob benefits from neuroscience theories and tools can advance and improve robot functions. In addition, CogRob methods and architectures can create artificial experimental settings to test variables in isolation [8]. To better incorporate the potential of CogRob methods and architectures for NeuroIS, we build on the proposed 2×2 matrix by vom Brocke and colleagues [4]. The model establishes how neuroscience theories and tools can help improve the design and evaluation of information systems within the DSR cycle. We expand this 2×2 by a $2 \times 2(+2)$ matrix, positioning CogRob methods and architectures as a valuable addition to designing and improving user-centric information systems (Fig. 2).

The expanded framework aims to demonstrate that CogRob can benefit NeuroIS design science research by complementing the theories and tools of neuroscience through computerized modeling. This capability expands the previously proposed three neuroscience application strategies in IS design science research by a CogRob dimension. The tentative application strategies for CogRob complement NeuroIS design science research as follows.

Strategy 1: *Use of neuroscience theories and cognitive robotic methods to build and evaluate user-centric IT artifacts.*

A study by Baima and Luna Colombini [29] showed that a robot could learn object affordances by interacting only with tactile sensors, thus simulating how blind users learn to interact with an unknown artifact. However, the architecture's adaptability allows the sensor type to change from tactile to vision or to have both as input, enabling the isolation of the evaluation and building of the IT artifact considering different scenarios. Such an approach would allow researchers to study variables more in-depth and use this knowledge to improve the design of artifacts regarding user-friendliness, productivity, and user experience.

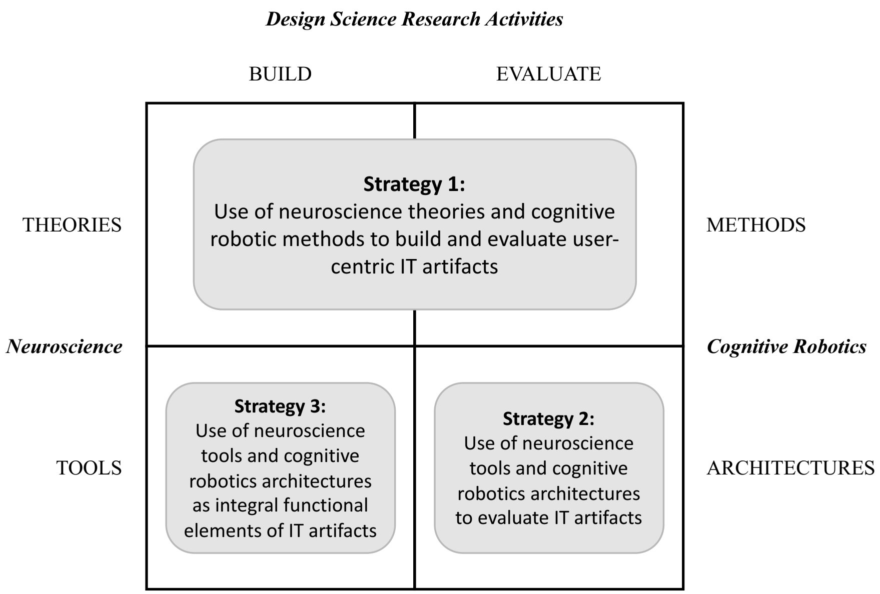


Fig. 2 The expanded NeuroIS Design Science Research Framework by CogRob methods and models for novel approaches to building and evaluating information systems (adapted from Brocke et al. [4])

Strategy 2: *Use of neuroscience tools and cognitive robotics architectures to evaluate IT artifacts.*

Researchers can isolate specific modules of an agent and observe its behavior under different conditions to make accurate comparisons between approaches. Researchers, for instance, can investigate the impact of two motivational mechanisms—e.g., needs and pleasure—on a simulated autonomous agent's decision-making behavior as part of its cognitive architecture [28]. This approach allows for the simulated isolation of the users' emotions in behavioral responses while interacting with applications and artifacts. Results can inform the redesign of the artifact in the simulation before the improved artifact is tested with the participants.

Strategy 3: *Use of neuroscience tools and cognitive robotics architectures as integral functional elements of IT artifacts.*

Incorporating cognitive architectural modules into chatbots, such as chatGPT [30], can enable self-adjustment and rebalancing of cognitive modules based on user feedback, either predicted (sentiment analysis) or user informed (through a "like"-button), allowing for the testing of new ideas in business models and their impact on user acceptance and trust. For instance, a study using this CogRob module demonstrated interactional justice's impact on the service recovery [31]. This approach allows NeuroIS researchers to create new paradigms and test human responses to behaviors while interacting with humans or other robots through IT artifacts.

5 Discussion

As an emerging field, Cognitive Robotics provides many interesting impulses for NeuroIS. Current benefits range from creating specific testing environments and testing variables in isolation to assessing user behavior and improving system designs based on user behavior. Moreover, machine learning algorithms in CogRob, combined with CogRob architectures and models, may help develop more practical models for brain-computer interactions. This interplay can adjust a user interface in near real-time, which may improve attitudes, performance, productivity, and well-being.

Despite the potential benefits of CogRob, it is challenging to integrate this emerging field into NeuroIS research and design. As an emerging field, CogRob requires a more explicit framework and definition to differentiate its approach from common machine learning and artificial intelligence frameworks [21, 32]. In particular, the term 'learning' is often used in the context of machine learning, which can be misleading regarding the true abilities of what machines are learning. It also needs to be more evident where CogRob can be used and to what extent. When it comes to, for instance, the testing of variables in a CogRob environment instead of an experiment with human participants, the current hardware limitations should be considered. Computers reach up to 500 processing cores, while the human brain processes information on billions of neurons in parallel. The result may not quickly transfer to the human organism [33].

6 Conclusion

NeuroIS and Cognitive Robotics are rapidly growing and interdisciplinary fields. Combined, they may significantly enhance our understanding of human behavior and intelligent systems. Building on vom Brocke and colleagues' work [4], we expand the 2×2 matrix on integrating neurobiology into the design science research cycle by CogRob methods and architectures. While preparing the $2 \times 2(+2)$ matrix, we introduce possible application areas at the intersection of CogRob and NeuroIS, highlighting how each discipline can cross-fertilize. Overall, cross-fertilization can lead to more effective and user-friendly cognitive architectures, aligning IT artifacts with users' perceptual and information-processing mechanisms, ultimately improving agent behaviors.

Acknowledgements This research was funded in part by the Luxembourg National Research Fund (FNR), in the FiReSpARX Project grant reference 14783405. Moreover, we would like to acknowledge the support of the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. For the purpose of open access, and in fulfilment of the obligations arising from the grant agreement, the authors have applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

References

1. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*, 9–24. https://doi.org/10.1080/0960085X.2019.1708218
2. Dimoka, A., Pavlou, P. A., & Davis, F. D. (2010). Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*. https://doi.org/10.1287/isre.1100.0284
3. Loos, P., Riedl, R., Müller-Putz, G. R., vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering, 2*, 395–401. https://doi.org/10.1007/s12599-010-0130-8
4. Brocke, J. V., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems, 53*, 1–13. https://doi.org/10.1080/08874417.2013.11645627
5. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain*. Springer.
6. Markram, H. (2013). Seven challenges for neuroscience. *Functional Neurology, 28*, 145–151.
7. Gravish, N., & Lauder, G. V. (2018). Robotics-inspired biology. *Journal of Experimental Biology, 221*, jeb138438. https://doi.org/10.1242/jeb.138438
8. Datteri, E., Chaminade, T., & Romano, D. (2022). Going beyond the "Synthetic Method": New paradigms cross-fertilizing robotics and cognitive neuroscience. *Frontiers in Psychology, 13*, 819042. https://doi.org/10.3389/fpsyg.2022.819042
9. Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B, 371*, 20150375. https://doi.org/10.1098/rstb.2015.0375
10. Christaller, T. (1999). Cognitive robotics: A new approach to artificial intelligence. *Artificial Life Robotics, 3*, 221–224. https://doi.org/10.1007/BF02481184
11. Holland, J., Kingston, L., McCarthy, C., Armstrong, E., O'Dwyer, P., Merz, F., & McConnell, M. (2021). Service robots in the healthcare sector. *Robotics, 10*, 47. https://doi.org/10.3390/robotics10010047
12. Angel, F.-A., Khanh, P., & Ou, M. (2013). *A review of robotics technologies for on-orbit services*. Defense Technical Information Center, 8725 John J. Kingman Road, Fort Belvoir, VA 22060–6218.
13. Riek, L. D. (2017). Healthcare robotics. *Communications of the ACM, 60*, 68–78. https://doi.org/10.1145/3127874
14. Ishii, K. (2006). *Cognitive robotics to understand human beings*. NISTEP Science & Technology Foresight Center.
15. Floreano, D., Ijspeert, A. J., & Schaal, S. (2014). Robotics and neuroscience. *Current Biology, 24*, R910–R920. https://doi.org/10.1016/j.cub.2014.07.058
16. Samani, H. (ed.). (2015). *Cognitive robotics*. CRC Press. https://doi.org/10.1201/b19171
17. Lange, D. (2019). Cognitive robotics: Making robots sense, understand, and interact. *Computer, 52*, 39–44. https://doi.org/10.1109/MC.2019.2942579
18. Feigenbaum, E. A. (1959). An information processing theory of verbal learning.
19. Anderson, J. R., & Bower, G. H. (2014). *Human associative memory*. Psychology Press. https://doi.org/10.4324/9781315802886
20. Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.
21. Krichmar, J. L. (2012). Design principles for biologically inspired cognitive robotics. *Biologically Inspired Cognitive Architectures, 1*, 73–81. https://doi.org/10.1016/j.bica.2012.04.003
22. Browne, W., Kawamura, K., Krichmar, J., Harwin, W., & Wagatsuma, H. (2009). Cognitive robotics: New insights into robot and human intelligence by reverse engineering brain functions [from the Guest Editors]. *IEEE Robotics & Automation Magazine, 16*, 17–18. https://doi.org/10.1109/MRA.2009.933617

23. Wang, Y. (2015). Cognitive learning methodologies for brain-inspired cognitive robotics. *International Journal of Cognitive Informatics and Natural Intelligence, 9*, 37–54. https://doi.org/10.4018/IJCINI.2015040103
24. Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*, 17–94. https://doi.org/10.1007/s10462-018-9646-y
25. Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation, 11*, 151–180. https://doi.org/10.1109/TEVC.2006.890274
26. Bar-Cohen, Y., & Breazeal, C. (2003). Biologically inspired intelligent robots. In *Smart Structures and Materials 2003: Electroactive Polymer Actuators and Devices (EAPAD)* (pp. 14–20). SPIE. https://doi.org/10.1117/12.484379
27. Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/A:1022193728205
28. Berto, L., Costa, P., Simões, A., Gudwin, R., & Colombini, E. (2023). *Learning goal-based movement via motivational-based models in cognitive mobile robots*. http://arxiv.org/abs/2302.09759, https://doi.org/10.48550/arXiv.2302.09759
29. Baima, R. L., & Luna Colombini, E. (2021) Modeling object's affordances via reward functions. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2183–2190). https://doi.org/10.1109/SMC52423.2021.9658915.
30. Luan, L., Lin, X., & Li, W. (2023). *Exploring the cognitive dynamics of artificial intelligence in the post-COVID-19 and learning 3.0 era: A case study of ChatGPT*. http://arxiv.org/abs/2302.04818, https://doi.org/10.48550/arXiv.2302.04818
31. Lajante, M., Remisch, D., & Dorofeev, N. (2023). Can robots recover a service using interactional justice as employees do? A literature review-based assessment. *Service Business, 17*, 315–357. https://doi.org/10.1007/s11628-023-00525-z
32. Ciria, A., Schillaci, G., Pezzulo, G., Hafner, V. V., & Lara, B. (2021). Predictive processing in cognitive robotics: A review. *Neural Computation, 33*, 1402–1432. https://doi.org/10.1162/neco_a_01383
33. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., & Yoshida, C. (2009). Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development, 1*, 12–34. https://doi.org/10.1109/tamd.2009.2021702

The Relationship Between Mental Effort and Social Value Orientation in Resource Allocation Problems

Dor Mizrahi, Ilan Laufer, and Inon Zuckerman

Abstract Resource allocation tasks are a central focus of game theory, where a player must allocate a limited set of resources among himself and others. Resource allocation tasks in game theory provide a framework for analyzing strategic decision-making in various economic, political, and social contexts. The Social Value Orientation (SVO) index is a concept used in social psychology to describe the degree to which individuals prioritize their interests over the interests of others in social dilemmas. In resource allocation tasks, the SVO may be used as an indicator to evaluate player's behavior. In this study, an experimental set-up was built to examine the relationship between the player's SVO index and his mental effort, which is measured by evaluating the Theta to Alpha ratio based on an EEG measurement. The results show a significant linear relationship between the player's SVO value and its mental effort. That is, the smaller the SVO value (a more competitive player), the greater the mental effort he invests in the resource allocation task.

Keywords EEG · Theta/Alpha Ratio · Social Value Orientation · Resource Allocation

1 Introduction

In game theory, resource allocation tasks (e.g. [1, 2]) refer to situations where a player must allocate a limited set of resources among himself and others. These resources can range from physical commodities like money to intangible assets like

D. Mizrahi (✉) · I. Laufer · I. Zuckerman
Department of Industrial Engineering and Management, Ariel University, Ariel, Israel
e-mail: dor.mizrahi1@msmail.ariel.ac.il

I. Laufer
e-mail: ilanl@ariel.ac.il

I. Zuckerman
e-mail: inonzu@ariel.ac.il

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_17

time or attention. The goal of each player is typically to maximize its own utility [3], which can be defined in several different ways depending on the game being played. Resource allocation tasks exists in many forms in the game theory literature: ranging from simple two-player games like the ultimatum game [4], where one player proposes a division of resources and the other player can either accept or reject the offer, to more complex games involving many players and multiple rounds of play (e.g. [5]). In some cases, the allocation of resources may be fixed, and players must compete to see who can claim the most valuable share of the resources. In other cases, players may be able to negotiate with one another to divide the resources in a mutually beneficial way [6].

One way to evaluate players' utility function via the Social Value Orientation (SVO) index [7, 8]. SVO is a concept used in social psychology to describe the degree to which individuals prioritize their interests over the interests of others in social dilemmas. SVO can be thought of as a continuum, with individuals at one end exhibiting a "prosocial" orientation, in which they prioritize the collective good and seek to maximize the benefits to all parties involved, while individuals at the other end exhibit a more "individualistic" orientation, in which they prioritize their self-interest and seek to maximize their gains, even at the expense of others. Individuals with a more prosocial SVO tend to exhibit greater trust, cooperation, and altruism. In comparison, those with a more individualistic SVO tend to exhibit greater competition and selfishness. Today, the most reliable way to measure the SVO value is through the "slider" questionnaire [7], which includes six questions about resource distribution. The result of the questionnaire is a number on a continuous scale that describes the player's preferences, where an SVO value of 0 represents an individualist player, and a value of 45 represents a prosocial player.

Previous research has shown that the Theta to Alpha Ratio (TAR) brainwaves is a good estimator of mental effort (e.g. [9, 10]), or the level of cognitive resources an individual is expending in a given task. Theta waves are associated with low-frequency brain activity that is typically observed during states of relaxation, daydreaming, or light sleep. In contrast, alpha waves are associated with higher-frequency brain activity typically observed during wakeful mental effort. Overall, the ratio of theta to alpha waves provides a valuable measure of mental effort, reflecting the underlying neural processes associated with cognitive engagement and attention. A high TAR value indicates low mental effort, and a low value indicates high mental effort. Following a previous study [11] which showed that there is a difference in the distribution of the TAR values for binary prosocial or individualistic labels, in this study, we would like to check whether there is a significant relationship between the *continuous SVO index* and the distribution of the TAR values. Our assumption in this study, following the results of previous studies (e.g. [2, 11, 12]), is that by using the SVO slider index [7], which is an extension of the SVO ring [13] index which generates a continuous SVO index (i.e., SVO angle) instead of a categorical index (i.e., prosocial or individualistic), we will be able to show a significant linear or polynomial relationship to electrophysiological indices.

2 Experimental Design

The study comprised the following stages. First, participants received an explanation regarding the overarching aim of the study and were given instructions regarding the experimental procedure and the interface of the application. Next, we have measured the SVO of the player using the "slider" method [7] and categorized it as "prosocial" or "individualistic". Then, each player was presented 18 resource allocation questions. Each such question included two resource-sharing options which should be divided between himself and another participant unknown in the experiment (Fig. 1 shows the resource allocation application screen). The number of points presented at the upper part of the screen is for the main player and the number of points at the lower part is for the other player. The player must choose within 8 s otherwise this will get no points. The position of the options on the screen was randomized. One option was the prosocial option (maximizing the joint profit of the two players), and the other an individualistic option (maximizing the personal profit of the player). Between the games a stand-by screen appeared for 1 s.

The participants were 10 students from the university that were enrolled in one of the courses on campus (right-handed, mean age = ~25, SD = 3). The EEG Data was recorded by a 16-channel g.USBAMP bio signal amplifier (g.tec, Austria) at a sampling frequency of 512 Hz, with 16 active electrodes based on the international 10–20 system. Recording was done by the OpenVibe [14] recording software. Impedance of all electrodes was kept below the threshold of 5 K [ohm] during all recording sessions.

We assessed the mental effort in each epoch using the Theta/Alpha ratio (TAR). The TAR measure of mental effort is based on the hypothesis that an increase in workload is associated with an increase in theta power with a simultaneous decrease in alpha power (e.g. [9, 15]). In terms of topographic distribution, previous studies have shown that workload decreased alpha power at parietal regions and increased theta power at frontal regions (e.g. [16]). Hence, in this study we have focused on the analysis of the of frontal and prefrontal cluster electrodes (Fp1, F7, Fp2, F8, F3, and F4). For each epoch we have calculated the accumulated mental effort [17], by

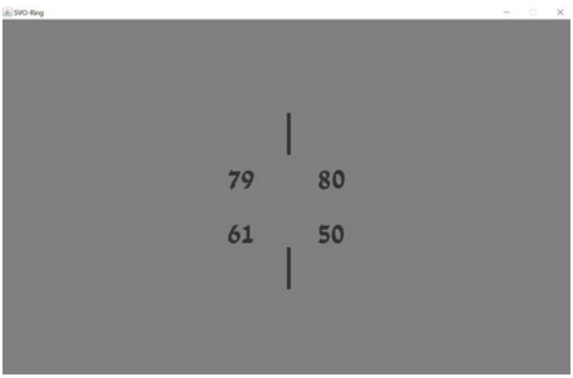


Fig. 1 Resource allocation application screen

calculating the energy ratio between theta and alpha bands for each participant on each single epoch.

3 Data Processing and Analysis

To improve the signal-to-noise ratio and reduce the effect of artifacts created during the EEG recording (e.g., eye movements, muscle activity, and electrical interference) before extracting the TAR value we have implemented data preprocessing pipeline, as was done in previous studies (e.g. [2, 10, 11]). The pipeline used a combined filter (band pass combined with a notch filter) followed by a re-reference scheme to an average reference, and decomposed using independent component analysis (ICA) [18]. To end the process the EEG signal than down sampled to 64 Hz following a baseline correction. Data was analyzed on a 1-s epoch window from the onset of each game.

Next, we estimated the intensity of the cognitive workload in each epoch using the TAR index. First, we have calculated the energy in the Theta and Alpha bands by using the Discrete Wavelet Transform (DWT) [19]. The DWT divides a signal into multiple frequency bands, each representing a different level of detail or approximation. This allows for a more efficient representation of the signal compared to other transforms, such as the Fourier transform, which represent a signal in terms of its frequency components only [19]. In this research, we used a 3-level DWT, which can extract the theta and alpha band (e.g. [2]) which are required to calculate the TAR index. The output of the DWT (i.e., the power of theta and alpha bands) is a signal which variant over the epoch time. To calculate the TAR for the entire segment we averaged the power of the whole epoch, (Eq. 1), and divided them to calculate the TAR index of the current epoch.

$$P_x = \frac{1}{T}\sum_{t=1}^{T} x^2(t) \tag{1}$$

We analyzed the distribution of 180 TAR values (18 epochs per player of 10 different players) that were extracted from the EEG segments according to their corresponding SVO value. We used a one-dimensional linear regression model to verify whether there is a relationship between the two variables, SVO and TAR, which represent a mental effort in resource allocation problems. The regression model showed that SVO significantly predicted TAR in resource allocation games. SVO also explained a significant proportion of variance in depression scores, $R^2 = 0.68716$, $p < 0.05$. Visualization of the average TAR values of each player according to his corresponding SVO index alongside the regression is presented in Fig. 2.

The TAR ratio represents the mental workload that the player invests during the task. The greater the workload the higher the TAR index should be (simultaneous change in two indices—decreased alpha and increased theta). These results,

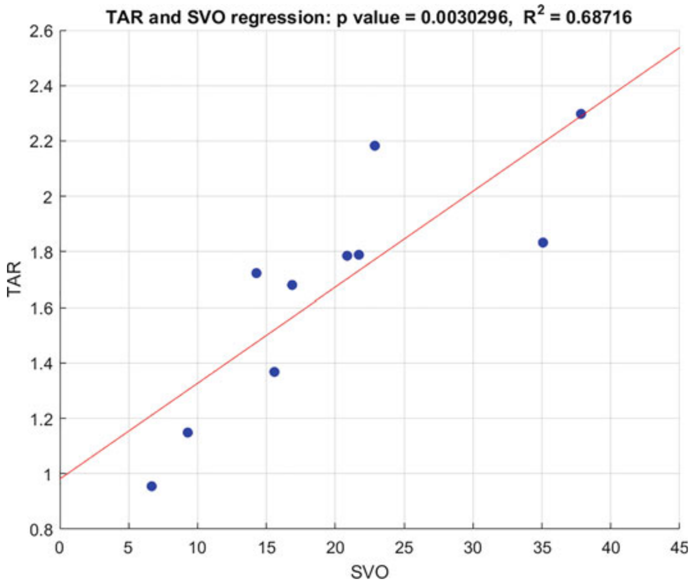


Fig. 2 TAR as function of SVO profiles

which are based on electrophysiological measurements, show that humans who are more prosocial according to the SVO theory (i.e., higher SVO index) invest a larger cognitive workload than their individualistic counterparts.

4 Conclusions and Future Work

This research presents for the first time the significant and positive relationship between the TAR, which is an electrophysiological marker of mental effort, and SVO in the context of resource allocation tasks. Specifically, this finding was demonstrated with the use of a prefrontal and frontal cluster of electrodes. This result corroborates previous research showing that SVO profiles may affect the strategic behavior of players [12], and that different behavioral strategies and indices may be accompanied by electrophysiological changes (e.g. [16, 20]). Following the results obtained in this study, there are many avenues for future research. First, it will be possible to examine the effect of the structure of the questionnaire, such as the absolute number of resources or the difference between the various options, on the distribution of electrophysiological indices or on the activity of specific regions in the brain. Second, according to [21] mental work load also affected parietal brain areas. It will be interesting to explore the effect of these areas on the TAR distribution and the correlation to the prefrontal and frontal areas. Third, it would be interesting to examine the distribution of electrophysiological results depending on the demographics and gender of the subject. Finally, previous studies have shown that other measures such culture [22, 23] and loss-aversion [24] may affect human behavior in decision making scenarios. It will be interesting to see if the TAR is also correlated with the abovementioned measures.

References

1. Nezarat, A., & Dastghaibifard, G. H. (2015). Efficient nash equilibrium resource allocation based on game theory mechanism in cloud computing by using auction. *PLoS One, 10*.
2. Mizrahi, D., Zuckerman, I., & Laufer, I. (2023). The effect of social value orientation on theta to alpha ratio in resource allocation games. *Information, 14*, 146.
3. Marden, J. R., & Roughgarden, T. (2014). Generalized efficiency bounds in distributed resource allocation. *IEEE Transactions on Automatic Control, 59*, 571–584.
4. Croson, R. T. (1996). Information in ultimatum games: An experimental study. *Journal of Economic Behavior & Organization, 30*, 197–212.
5. Slembeck, T. (1999). Reputations and fairness in bargaining-experimental evidence from a repeated ultimatum game with fixed opponents.
6. Lee, M., Lucas, G., & Gratch, J. (2021). Comparing mind perception in strategic exchanges: Human-agent negotiation, dictator and ultimatum games. *J. Multimodal User Interfaces, 15*, 201–214.
7. Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision making, 6*, 771–781.
8. Mizrahi, D., Laufer, I., & Zuckerman, I. (2021). The effect of expected revenue proportion and social value orientation index on players' behavior in divergent interest tacit coordination games. In: *International Conference on Brain Informatics* (pp. 25–34). Springer.
9. Fernandez Rojas, R., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., & Abbass, H. (2020). Electroencephalographic workload indicators during teleoperation of an unmanned aerial vehicle shepherding a swarm of unmanned ground vehicles in contested environments. *Frontiers in Neuroscience, 14*, 1–15.
10. Mizrahi, D., Zuckerman, I., & Laufer, I. (2022). Electrophysiological features to aid in the construction of predictive models of human-agent collaboration in smart environments. *Sensors, 22*, 6526.
11. Mizrahi, D., Zuckerman, I., & Laufer, I. (2022). The effect of SVO category on theta/alpha ratio distribution in resource allocation tasks. In: *International Conference on Brain Informatics.*
12. Mizrahi, D., Laufer, I., & Zuckerman, I. (2021). Predicting focal point solution in divergent interest tacit coordination games. *Journal of Experimental & Theoretical Artificial Intelligence,* 1–21.
13. Liebrand, W. B., & Mcclntock, C. G. (1988). The ring measure of social values : A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality, 2*, 217–230.
14. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., & Le´cuyer, A. (2010). Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence: Teleoperators and Virtual Environments, 19*, 35–53.
15. Stipacek, A., Grabner, R. H., Neuper, C., Fink, A., & Neubauer, A. (2013). Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. *Neuroscience Letters, 353*, 193–196.
16. Laufer, I., Mizrahi, D., & Zuckerman, I. (2022). An electrophysiological model for assessing cognitive load in tacit coordination games. *Sensors, 22*, 477.

17. Bagyaraj, S., Ravindran, G., & Shenbaga Devi, S. (2014). Analysis of spectral features of EEG during four different cognitive tasks. *International Journal of Engineering & Technology, 6*, 725–734.
18. Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*, 411–430.
19. Jensen, A., & la Cour-Harbo, A. (2001). *Ripples in mathematics: The discrete wavelet transform*. Springer Science & Business Media.
20. Mizrahi, D., Laufer, I., & Zuckerman, I. (2021). Level-K Classification from EEG signals using transfer learning. *Sensors., 21*, 7908.
21. Zhang, P., Wang, X., Chen, J., You, W., & Zhang, W. (2019). Spectral and temporal feature learning with two-stream neural networks for mental workload assessment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27*, 1149–1159.
22. Mizrahi, D., Laufer, I., & Zuckerman, I. (2020). Collectivism-individualism: Strategic behavior in tacit coordination games. *PLoS One, 15*.
23. Cox, T. H., Lobel, S. A., & Mcleod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal, 34*, 827–847.
24. Mizrahi, D., Laufer, I., & Zuckerman, I. (2020). The Effect of loss-aversion on strategic behaviour of players in divergent interest tacit coordination games. In: *International Conference on Brain Informatics* (pp. 41–49). Springer.

It's Not Me, It's You: Breaking Up with ICT to Cope with Techno-Unreliability. A Proposed fMRI Experiment

Nina Melanie Mølgaard and Mirja Hubert

Abstract This study will investigate the neural correlates associated with the transactional process of technostress. Specifically, we will focus on the automatic and implicit processes of the primary appraisal process of techno-unreliability in the form of delayed system response time (SRT) and system breakdown, as well as the neural activation pattern of the strain experienced as a result hereof, leading to the secondary appraisal process of coping responses (e.g., attentional and behavioral (dis)engagement) with (dis)continued use of information and communication technology (ICT) as a potential outcome; all in the specific context of a consumer decision making process facilitated by ICT using functional magnetic resonance imaging (fMRI). With this proposed research we aim to contribute to the current technostress literature by providing new insights on the neural correlates of technostress, specifically the primary and secondary appraisal processes.

Keywords Techno-unreliability · Technostress · Appraisal processes · Discontinuance · Coping · Functional magnetic resonance imaging (fMRI) · NeuroIS

1 Introduction

Information and communication technology (ICT) continues to facilitate the ever-evolving digitalization of modern life and with over 2.14 billion people worldwide expected to purchase goods and services using ICT in 2021, and the number of online customers is expected to continue to rise [1]. Despite the euphoria about the promises of digital transformation, these developments also have a "dark side" and can yield

N. M. Mølgaard (✉) · M. Hubert
Department of Management, Aarhus University, Aarhus, Denmark
e-mail: nmmn@mgmt.au.dk

M. Hubert
e-mail: mirja.hubert@mgmt.au.dk

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_18

distressing effects for both consumers and companies. In this research-in-progress we want to propose the investigation of neural correlates associated with the process of technostress as one specific negative phenomenon associated with technology use. Technostress can be defined as "*stress that results from both the use of information and communication technologies (ICTs) and the pervasiveness and expectations of ICT use in society in general*" [2] (p. 376). It is often conceptualized as a transactional process, that emerges from the interaction between the individual and the technological environment, which evolves when the subjectively perceived environmental and technological demands exceed the resources of an individual [2–5]. According to this, our conceptualization identifies technostress as a transactional process that includes potential techno-stressors (here: techno-unreliability), associated psychological responses (i.e., individual strain and individual coping) translating into behavioral outcomes (here: (dis)continuance behavior) relevant for information systems (IS) research [2, 6, 7] (Fig. 1). To better understand the process of technostress, going beyond just the stressor component, it is important to investigate related implicit and automatic primary and secondary appraisal processes, which are difficult to capture with survey designs. During the primary appraisal process the individual assesses the extent to which the interaction with a technology have the potential to yield individual strain. The secondary appraisal evaluates different coping responses and resources in an effort to prevent harm or to improve the outcome [7]. Even though many articles investigating the effect of technostressors on different behavioral and strategic outcomes already exist, only very few include implicit and automatic processes in their studies (e.g., [8]) and the understanding of the decision-process as well as underlying neurological processes of ICT discontinuance behavior as an outcome of a coping response resulting from experienced strain remain scarce [9, 10]. Therefore, this study seeks to add to the scarce understanding of the neural processes, more specifically regarding the brain, underlying the primary and secondary appraisal processes included in the technostress process. On the basis hereof, we propose the following research question: *What are the underlying neural processes associated with (I) the primary appraisal of potential technostressors (here: perceived techno-unreliability) that potentially generates individual strain (e.g., negative psychological response), as well as (II) the secondary appraisal regarding the selection of a coping response (e.g., attentional & behavioral disengagement) that leads to a specific behavioral outcome (e.g., discontinuance behavior)?*

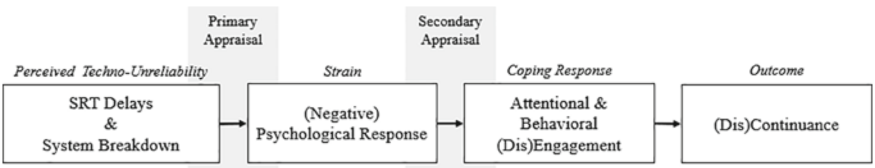


Fig. 1 Proposed research framework

2 Theoretical Background

The transactional model of stress (TMS) by [11] is the main theoretical framework used in the IS literature to explain *technostress* [4, 5, 10]. Based on the TMS framework, technostress is understood as a process including (1) techno-stressors (e.g., techno-unreliability) that lead to (2) strain on the individual (e.g., negative psychological response), therefore requiring an appropriate (3) coping response (e.g., disengagement) followed by (4) a coping outcome (e.g., discontinued use) [9]. For technostress to be experienced, the individual will need to appraise the interaction with technology as relevant and stressful followed by the choice of a sufficient coping response to reduce the experienced strain [2]. ICTs can evoke stress in a complex way and [4] identified five essential stressors associated with technostress, which they called "technostress creators" [4, 7, 12]. These technostress creators are defined as "factors that create stress from the use of ICTs" [4] (p. 417) and have been widely used in technostress research—often to determine the level of technostress [2, 4, 12–18]. Some studies discuss if techno-unreliability (e.g., system response time (SRT) delays) should be considered a separate technostressor, and recent research has conceptualized it as part of techno-overload [9]. Studies show that individual strain (i.e., the negative psychological outcome of encountering a technostressor) can make individuals discontinue using technology altogether [19–24].

This study seeks to investigate the coping outcome preceded by a coping response, more specifically the decision to either (1) continue, (2) temporarily discontinue (i.e., take a break), or (3) permanently discontinue using the ICT perceived to be unreliable and/or inconsistent (i.e., *perceived techno-unreliability* [6, 25, 26]) as well as the underlying neural processes involved herein. Techno-unreliability will be manipulated in two ways based on previous literature: (1) SRT delays [10] and (2) system breakdowns causing loss of task progression [27] in an online consumer decision making context. See Fig. 1 for our proposed research framework. On a behavioral level we hypothesize:

- **H1**: Higher *perceived techno-unreliability* will yield higher *strain* and set in motion *coping responses* (here: disengagement/discontinuance behavior).

  Currently, the studies investigating technostress from a neurobiological perspective is scarce, especially investigations regarding the neural processes within the brain. However, there are investigations that are focused on negative emotions and stress outside of the technostress literature that has used methods such as functional magnetic resonance imaging (fMRI) to better explain the underlying neural processes hereof. For example, the generation, expression, and experience of negative emotion (e.g., stress) has been related mainly to activity in the amygdala and the insula [28, 29], whereas coping with the experienced negative emotions have been seen to rely on activation in mainly the frontal brain regions, such as the orbitofrontal cortex (OFC), dorsolateral prefrontal cortex (dlPFC), dorsal medial prefrontal cortex (dmPFC), ventrolateral prefrontal cortex (vlPFC), and anterior cingulate cortex (ACC) [29], as well as the strength of the connectivity between the mentioned frontal brain regions and the amygdala [29]. We therefore propose the following hypotheses:

- **H2**: The appraisal of *perceived techno-unreliability* will yield activity changes in regions associated with stress and negative emotions such as the *insula* and the *amygdala.*
- **H3**: The appraisal of coping responses (e.g., attentional, and behavioral (dis)engagement) following *perceived techno-unreliability* will yield activity changes in regions associated with coping processes such as the *OFC*, *dlPFC*, *dmPFC*, *vlPFC*, and *ACC.*
- **H4**: The outcome (i.e., the decision to (dis)continue) following the coping response will yield activity changes in regions associated with coping processes such as the *OFC*, *dlPFC*, *dmPFC*, *vlPFC*, and *ACC.*
- **H5**: Increased connectivity between the frontal brain regions (e.g., the *OFC*, *dlPFC*, *dmPFC*, *vlPFC*, and *ACC*) and the *amygdala* will be linked to an increased ability to self-regulate the experienced strain (e.g., negative emotion) following the *perceived techno-unreliability.*

3 Methodology

Experimental Design

Upon arrival, participants will be asked to sign a consent form. To incentivize participants, they will be told that their compensation will be based on their level of engagement with the task. They will then receive training in the task that they are asked to perform in the fMRI scanner, in order to make sure that they have understood the assignment before entering. Throughout the task, participants will answer questions using a 4-button fiber optic response pad inside the scanner. This setup will be emulated in the practice task performed outside the scanner in order for participants to get familiar with the response pad setup before entering the scanner. None of the trials will exceed four possible answers in order to make sure that each answer corresponds with an assigned button on the response pad.

The experimental design will follow a block design and will be divided into three blocks (Fig. 2). The first block will serve as a baseline measure for the neural activity captured during the subsequent blocks. Each block will consist of the same consumer decision making task (see Sect. 3.2), although blocks 2 and 3 will include technostress manipulations (see Sect. 3.3) in an attempt to induce techno-uncertainty in the form of SRT delays introduced in block 2 and system breakdown introduced in block 3. The fMRI data will be analyzed both using block and event-related methods of analyses. The data from block 2 and 3 will be contrasted to block 1 using block analysis. Additionally, an event-related analysis will be performed on data from block 2 where the trials including SRT delays will be contrasted to the trials without these manipulations.

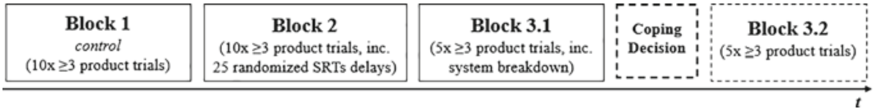


Fig. 2 Overall experimental protocol

Consumer Decision Making Task

Each block will start with participants viewing a list of 10 grocery items from different product categories (e.g., pasta, bread, soda) resembling a grocery list. Participants will have to add one product from each category to a digital shopping cart throughout the task. The shopping cart will be displayed in every trial in the upper right corner, counting the number of products put in the cart by the participant. Participants will be shown a minimum of three products per category, and every product will be shown for a minimum of three seconds, after which the participants will be able to press a button to continue to the next product or stop the information search, when they believe to have enough information to decide on the product they want to put in the cart. Participants will be allowed to browse the number of products that he/she desires, each category containing approx. 10 different products. If the participant chooses to stop browsing before all products within a category has been displayed, he/she will answer a question to why they chose to stop. If the participant browses all products within a category, they will be asked why they did *not* choose to stop browsing. The task is predominantly inspired by [30].

Then, participants will briefly view each of the products, that they have browsed through, before deciding which of the products they would like to add to their cart (Fig. 3). After adding the product to their cart, the number on the cart will increase by one. The choice as well as the reason for stopping the information search in each of the product categories is of no interest to this study, although the participants will not be aware of this fact.

Technostress Manipulations

During block 2, a spinning wheel in the middle of the screen will be shown the task for 3–9 s, simulating SRT delays [10, 31] to induce technostress. There will be 25 simulated SRT delays which will be counterbalanced throughout the block.

In block 3, a system breakdown will be introduced between product category five and six [27]. The first product of the sixth category will show for three seconds followed by a three second SRT delay. After the SRT delay, the screen will return although the shopping cart will show a zero. The loading wheel will appear once again; this time for nine seconds, followed by an error message, telling the participants that due to prolonged loading, the session has expired, and the participant will be asked to decide, if they would want to either (1) start over with the task (*continue*),

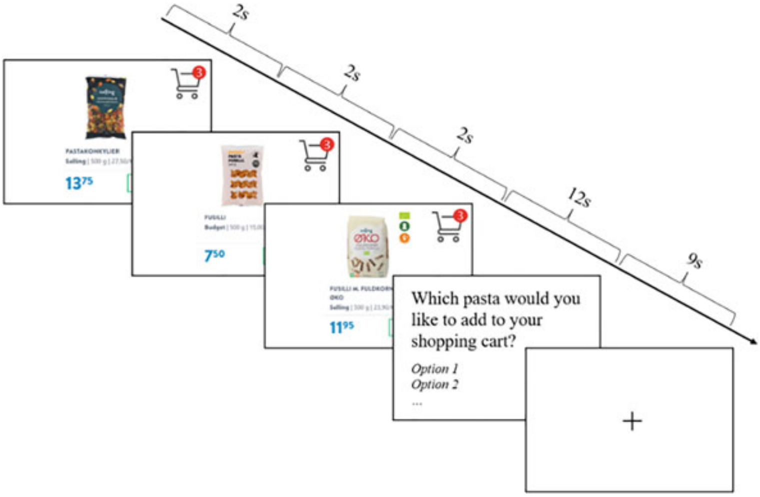


Fig. 3 Protocol for product decision making

(2) have a break before continuing (*temporarily discontinue*), or (3) stop the task (*discontinue*). Once they have answered, a blank screen with a fixation cross will be displayed for twelve seconds. The task will continue based on the participant's outcome choice. Participants continuing or temporarily discontinuing with the task will move on to block 3.2 (Fig. 2). The participants choosing to discontinue will be let out of the scanner. After participants have been let out of the scanner, they will be debriefed, compensated, and thanked for their participation.

Techno-unreliability in the form of SRT delays and system breakdowns was chosen as they have already been established as being able to produce a neurophysiological response in previous literature [10, 27, 31], providing a basis for adding brain imaging techniques to the existing literature, which to the knowledge of the authors, as not been done previously; particularly the underlying neural processes of the primary and secondary appraisal processes has not been investigated in relation to the technostress process using brain imaging techniques.

4 Limitations, Contributions and Future Work

With this proposed research we want to contribute to the current technostress literature by providing new insights on the neural correlates of technostress. This is in line with several studies that call for more experimental studies and propose the inclusion of physiological measurements in technostress research [8, 9, 26, 32]. A better understanding of the implicit and automatic processes associated with the interaction with technology is crucial for IS research, because it can help both users as well as companies to decrease technostress and avoid discontinuance behavior.

Due to the relatively simple experimental set-up needed for an fMRI study we cannot include the full scope of complexity of the technostress construct; this is a clear limitation of our proposed study. It would of course also be very interesting to further understand which effect other technostressors (e.g., techno-invasion) have regarding the evoked brain activations.

Our future work is aimed at conducting several experimental studies both inside and outside the scanner. It could be interesting to employ other physiological measurements such as hormonal level, heartrate, eye tracking or skin conductance as a supplement to the fMRI as well as other technostressors.

References

1. Coppola, D. (2021). *Number of digital buyers worldwide from 2014 to 2021.* [06/03/23]; Available from: https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/.
2. Fischer, T., & Riedl, R. (2017). Technostress research: A nurturing ground for measurment pluralism? *Communications of the Association for Information Systems, 40*(17), 375–401.
3. Fuglseth, A. M., & Sørebø, Ø. (2014). The effects of technostress within the context of employee use of ICT. *Computers in Human Behavior, 40*, 161–170.
4. Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433.
5. Zhao, X., Xia, Q. & Huang, W. (2020). Impact of technostress on productivity from the theoretical perspective of appraisal and coping processes. *Information & Management, 57*(8).
6. Califf, C. B., Sarker, S., & Sarker, S. (2020). The bright and the dark sides of technostress: A mixed-methods study involving healthcare IT. *MIS Quarterly, 44*(2), 809–856.
7. Tarafdar, M., Cooper, C. L., & Stich, J.-F. (2019). The technostress trifecta - techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal, 29*, 6–42.
8. Riedl, R. (2012). On the biology of technostress: literature review and research agenda. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 44*(1), 18–55.
9. Nastjuk, I., Trang, S., Grummebeck-Braamt, J.-V., Adam, M. T. P., & Tarafdar, M. (2023). Integrating and synthesising technostress research: A meta-analysis on technostress creators, outcomes, and is usage contexts. *European Journal of Information Systems,* 2154712.
10. Korosec-Serfaty, M., Riedl, R., Sénécal, S., & Léger, P.-M. (2022). Attentional and behavioral disengagement as coping responses to technostress and financial stress: An experiment based on psychophysiological, perceptual, and behavioral data. *Frontiers in Neuroscience, 16*, 883431.
11. Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of Personality, 1*(3), 141–169.
12. Tarafdar, M., Tu, Q., & Ragu-Nathan, T. S. (2014). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems, 27*(3), 303–334.
13. Tarafdar, M., Qiang, T., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The impact of technostress on role stress and productivity. *Journal of Management Information Systems, 24*(1), 301–328.
14. Shu, Q., Tu, Q., & Wang, K. (2011). The impact of computer self-efficacy and technology dependence on computer-related technostress: a social cognitive theory perspective. *International Journal of Human-Computer Interaction, 27*(10), 923–939.
15. Srivastava, S. C., Chandra, S., & Shirish, A. (2015). Technostress creators and job outcomes: Theorising the moderating influence of personality traits. *Information Systems Journal, 25*(4), 355–401.

16. Khan, A., Rehman, H., & Ur Rehman, S. (2013). An empirical analysis of correlation between technostress and job satisfaction: a case of KPK, Pakistan. *Pakistan Journal of Library and Information Science, 14*, 9–15.
17. Alam, M. A. (2016). Techno-stress and productivity: Survey evidence from the aviation industry. *Journal of Air Transport Management, 50*, 62–70.
18. Hwang, I., & Cha, O. (2018). Examining technostress creators and role stress as potential threats to employees' information security compliance. *Computers in Human Behavior, 81*, 282–293.
19. Dhir, M., & Midha, V. (2014). Overload, privacy settings, and discontinuation: A preliminary study of FaceBook users. In: *SIGHCI 2014 Proceedings.*
20. Kim, S., Park, H., & Choi, M. J. (2019). Negative impact of social network services based on stressor-stress-outcome: The role of experience of privacy violations. *Future Internet, 11*(6), 137.
21. Zhang, S., Zhao, L., Lu, Y., & Yang, J. (2016). Do you get tired of socializing? An empirical explanation of discontinuous usage behaviour in social network services. *Information & Management, 53*(7), 904–914.
22. Fu, S., Li, H., Liu, Y., Pirkkalainen, H., & Salo, M. (2020). Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload. *Information Processing & Management,* 57(6).
23. Chaouali, W. (2016). Once a user, always a user: Enablers and inhibitors of continuance intention of mobile social networking sites. *Telematics and Informatics, 33*(4), 1022–1033.
24. Luqman, A., Cao, X., Ali, A., Masood, A., & Yu, L. (2017). Empirical investigation of Facebook discontinues usage intentions based on SOR paradigm. *Computers in Human Behavior, 70*, 544–555.
25. Adam, M. T. P., Gimpel, H., Maedche, A., & Riedl, R. (2017). Design blueprint for stress-sensitive adaptive enterprise systems. *Business & Information Systems Engineering, 59*(4), 277–291.
26. Weinert, C., Maier, C., Laumer, S., & Weitzel, T. (2020). Technostress mitigation: An experimental study of social support during a computer freeze. *Zeitschrift für Betriebschaft, 90*(8), 1199–1249.
27. René Riedl, H. K., Aulnger, A., & Javor, A. (2012). Technostress from a neurobiological perspective: System breakdown increases the stress hormone cortisol in computer users. *Business & Information Systems Engineering,* 61–69.
28. McRae, K., Hughes, B., Chopra, S., Gabrieli, J. D. E., Gross, J. J., & Ochsner, K. N. (2010). The neural bases of distraction and reappraisal. *Journal of Cognitive Neuroscience, 22*(2), 248–262.
29. Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala-frontal activity during emotion regulation. *Social Cognitive and Affective Neuroscience, 2*(4), 303–312.
30. Browne, G. J., & Walden, E. A. (2021). Stopping information search: An fMRI investigation. *Decision Support Systems, 143*, 113498.
31. Riedl, R., Fischer, T., Xiao, B., & Nah, F. (2018). System response time as a stressor in a digital world: Literature review and theoretical model. *HCI in Business, Government, and Organizations, 10923*, 175–186.
32. Davis, F. D., Riedl, R., Brocke, J. v., Léger, P.-M., & Randolph, A. B. (2018). NeuroIS: A survey on the status of the field. *Information Systems and Neuroscience (NeuroIS Retreat 2018), 29*, 1–10.

Customer Decision-Making Processes Revisited: Insights from an Eye Tracking and ECG Study Using a Hidden Markov Model

Tobias Weiß, Lukas Merkl, and Jella Pfeiffer

Abstract Good timing is key for many activities in business and society. In the context of adaptive user assistance, it can work as door opener to further engage with the user. This paper presents a virtual commerce study which combines eye tracking, electrocardiography, and virtual reality with the goal to detect decision phases in two different purchase scenarios. We therefore collect objective sensor data in combination with subjective decision phase annotations. Shifts between decision phases are determined subjectively by the participants via retrospective video analysis. For decision phase recognition, we demonstrate how to use the neurophysiological sensor data to train a Hidden Markov Model with multivariate mixed Gaussian emission distributions and how to use it for inference. A main benefit of our approach is its lightweight character regarding both training and inference.

Keywords Customer behavior · Decision making · Eye tracking · Electrocardiography · Hidden Markov model · Gaussian mixture model · Machine learning · Virtual commerce · Virtual reality

1 Introduction

Approaching customers at the right time is crucial because it can significantly impact the interaction success [27]. Specifically, good timing can help to maximize engagement, build trust, and increase conversion rates [6]. However, to determine the right point in time to approach a customer requires profound understanding of the

T. Weiß (✉) · L. Merkl · J. Pfeiffer
Justus Liebig University Gießen, Gießen, Germany
e-mail: tobias.weiss@wirtschaft.uni-giessen.de

L. Merkl
e-mail: lukas.merkl@wirtschaft.uni-giessen.de

J. Pfeiffer
e-mail: jella.pfeiffer@wirtschaft.uni-giessen.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_19

target audience's behavior and preferences [8]. Advances in conversational agents and user assistance systems (UAS) often focus on the right information, introduce context awareness and improve interactivity [12, 17, 27]. Rather scarcely, previous research has investigated invocation timing based on neurophysiological indicators [16]. Within the ongoing transformation of the retail sector towards virtual commerce [3] and the rise of the metaverse idea [1], good invocation timing is one of the key components for a variety of information system (IS) artifacts. Decades ago, metaverse and virtual reality (VR) advocates already envisioned that a large fraction of daily life and therewith a large fraction of shopping activities transfers to virtual spaces [10, 25]. Today, this process gains momentum, as big tech companies introduce new hardware and applications with rigorous commitment. Latest VR headsets ship with eye and face tracking technology which fosters the potential and feasibility of neurophysiological IS and therefore turns them into a game changer. With a VR headset on their head, future customers wear a variety of sensors in proximity to the most reliable information source about their attitudes and moods. In this paper, we present our approach to integrate neuroscientific methods into virtual commerce IS. Our research question states as follows:

- **RQ**: Can we determine a good timing to approach customers in a virtual commerce scenario using eye-tracking and electrocardiography?

We report our insights gained from a study in which 50 participants had to make purchase decisions for either washing powder or 3D printers while wearing a head-mounted VR headset. We collected participants' eye tracking (ET) data, electrocardiography (ECG) data, and created a prediction model that can distinguish between different decision phases. Our insight can be used to inform a UAS or digital human agent when help is wanted. As model for decision phase recognition, we chose a combination of multivariate Gaussian Mixed Model and Hidden Markov Model (GMM-HMM). The benefit of our approach is its lightweight character in both training and inference. Thus, the presented GMM-HMM approach offers itself as good candidate to make it into soon-to-be released virtual commerce (and other) neurophysiological sensor based IS artifacts [28]. To the best of our knowledge, no study exists which applies machine learning approaches to differentiate between different decision phases using neurophysiological sensor data. Our research builds up on previous models but tries to apply a more generic inference method not solely dependent on product comparisons and re-dwells.

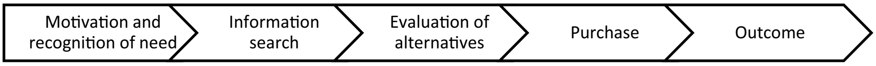


Fig. 1 The EKB model, dividing customer decision processes into five phases [5]

2 Method

Consumer Decision-Making

Several scholars in consumer behavior research suggested models to subdivide customer decision-making processes into different phases. Most studies support a phase theory which consists at least of an orientation and an evaluation phase. One prominent phase model is the five-stage Engel Kollat Blackwell (EKB) model [5], as shown in Fig. 1. The EKB model is still widely accepted [24] and frequently serves as basis for further adjustment to integrate specific aspects and research field dependent needs, such as modifications for an eye tracking study in VR.

In an eye tracking context, several other decision phase models were developed, e.g., by [7, 16, and 21]. These models subdivide decision processes into three phases—orientation, evaluation, and validation. The transition between different phases is based on simple rules, like re-fixations on products. The VR study in [16] pursued an on-the-fly attempt to determine the phases. Its authors used eye tracking data and identified the first comparison between two products as shift between orientation and evaluation. Furthermore, the shift between evaluation and verification was considered as the moment when the first product entered the shopping cart (We believe this is a questionable criterion because putting a product into the shopping cart signals a certain level of confidence).

For the right timing of user assistance, we consider the shift between orientation and evaluation as particularly interesting. We conjecture that help is most appreciated by customers after being within the evaluation phase for a certain offset duration. To verify this assumption empirically, self-reported desired help timings can be used. Knowing the phase of a decision process and the offset duration at least approximately, a UAS or sales representative can determine a good starting point to approach the customer.

Neurophysiological Data Collection in VR

The development of visual VR has a longer history than one might expect. For example, an early head mounted display (HMD) was already developed by Sutherland [26]. Commercial endeavors of big tech companies still focus on HMD development. For research, the latest HMD generation is particularly interesting because many models ship with integrated neurophysiological sensors, particularly ET [18]. ET is integrated because it can be used to optimize graphic card utilization via foveated rendering, a method which only renders the focused area in high detail while neglecting peripheral areas [14]. Recent research-grade HMDs include further sensors as ECG and electroencephalography (EEG). The integration of EEG into consumer-grade hardware seems rather unrealistic in the near and intermediate-term future as the sensor itself is expensive and the electrodes are relatively uncomfortable to wear. ECG measures a person's heart rate and is more likely to find its way into consumer devices. Another sensor, which is very likely to be included into future customer-grade HMDs, is photoplethysmography (PPG). PPG is a light-based sensor which can also be used to measure heart rate and corresponding metrics. Compared to ECG, PPG is cheaper, easier to attach (e.g., a forehead-sensor integrated in the HMD-cover), but less accurate. It is also imaginable to couple wearables with an HMD, particularly fitness watches, which already include ECG or PPG sensors. Overall, ET and ECG/PPG are the most likely sensors for future off-the-shelf HMDs. Thus, it makes sense to use gaze patterns, pupillometry, and heart rate as data sources for inference.

Hidden Markov Model

An HMM is a statistical model which describes a Markov process with a set of states between which it can transition [4, 20]. At each state, an HMM generates an observation or output symbol, which is associated with that state. Such observations generated by a state of the model are referred to as emissions. HMMs find application in a variety of disciplines [9, 11, 23]. To match the characteristics of our purchase decision scenario in the experimental VR setup, we use elements of both the classic EKB phase model [5] and the eye tracking model proposed by Russo et al. [21]. We begin with a memorization phase which corresponds to the motivation phase of the EKB model. During this phase, participants see purchase criteria on a blackboard and memorize them. The transition between memorization and the next phase is identified by a button press. For the subsequent phases, we use the phase labels orientation, evaluation and verification as proposed by Russo et al. [21]. However, we outline that the state transitions in our model have nothing in common with the originally proposed transitions which were based on specific gaze patterns. Instead, shifts to evaluation and verification were determined via self-reported timestamps given by the participants. Next, we adopt the purchase phase from the EKB model, as participants remained inside the VR scenario after confirming the purchase. Furthermore, an initial and terminal state are added as they are needed for computation. The corresponding HMM with flat prior transition probabilities is shown in Fig. 2.

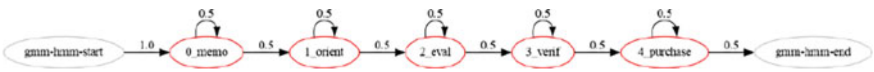


Fig. 2 GMM-HMM with flat prior transition probabilities

When the model transitions from one state to another, it refers to a (hidden) multivariate probability distribution which corresponds to the current input features. Internally, each state holds a multivariate Gaussian mixture distribution (what turns the model into a GMM-HMM), which is trained with ET and ECG features based on consecutive five second time windows. For each of these windows, our feature engineering pipeline creates 44 features which comprise 26 ET and 18 ECG features. ET features consist of statistical moments (mean, min, max, var) for blinks, fixations, fixation duration, pupil size, saccadic duration, and saccadic speed. ECG features are limited to the time domain, particularly the heart rate and its variability. Frequency domain related and non-linear ECG features are not considered because they would require longer window durations [19]. If participants indicate a state transition during such a window, the label for the subsequent and all following windows changes to the next state.

For real-time inference, the GMM-HMM can even be simplified to a GMM classifier which decides if the evaluation phase is reached or not. Features of a current observation can be shown to the model which maps them to the probability distribution and stochastically decides whether the evaluation phase is reached or not. If the evaluation phase is indicated several times in a row, the offset of approximately fifty seconds could be added and finally the UAS or digital human agent could approach the customer with a help offering.

3 Experiment

Participants

Our sample was collected from 50 participants (29 female, mostly students) with a mean age of 24.5 years (SD = 4.89). Only individuals with normal or corrected-to-normal vision via contact lenses were accepted since glasses would not fit into the HMD and not wearing them might confound the ET data. The participation compensation consisted of a fixed 10 Euro baseline plus a performance-based component. After arrival at the lab, participants signed a consent form. It ensured the participants' basic knowledge of the experiment procedure and informed them that the experiment complied with ethical standards. Further, it required them to grant permission to publish their data in anonymized form. For recruitment, we used the participant pool in our self-hosted online registration platform [2] and actively approached students on campus.

Procedure

We simulated customer purchase decisions in VR, collecting ET and ECG data. All virtual scenes were created using the Unity 2021.3 game engine. Participants entered our showroom using a Varjo VR 3 HMD with high-frequency ET capability (sampling rate of up to 200 Hz) and a display resolution of 2880 × 2720 pixels per eye. A bioPLUX device was used for ECG recording and captured signals with a sampling rate of 1000 Hz. Overall, the experiment followed a between-subjects design and included two different decision scenarios, one for 3D printers and one for washing powders (see Fig. 3). To create realistic shopping scenarios, we presented dedicated cover stories to both groups. Participants were then shown a list of purchase decision criteria they had to memorize. The end of memorization phase was triggered by the participants using a button press which hid the criteria and spawned the products. Then, they had the chance to gain one Euro in addition to their participation compensation if they matched a previously determined team decision. This monetary incentive helped to motivate the participants and increased the external validity of the experiment. Participants confirmed their purchase decision either by putting the product into a shopping cart or by clicking a purchase button. After making the purchase, participants left the VR environment and answered questions about their decision phases by means of a first-person video. This video showed a gaze dot which indicated their visual attention. Participants determined the moments when they shifted (1) from orientation to evaluation and (2) from evaluation to verification. For each of these phase shifts, they entered the timestamp in a web-based questionnaire form. Furthermore, participants reported their desired help timing for a digital human agent in the same manner as for the phase shifts.

4 Results

For our analysis, we used python 3.10 and the neurokit2 0.2.3 [13], pomegranate 0.4.0 [22], and scikit-learn 1.0.2 [15] packages.

We verified our conjecture regarding the desired help timing. As expected, help was most frequently desired after the shift from orientation to evaluation but before entering the verification phase. On average, the phase shift from orientation to evaluation was indicated after 100.2 s (SD = 79.8) and the shift from evaluation to verification was after 210 s (SD = 97.2). Participants reported the average desired help timing for a digital human agent after 148 s (SD = 115.8), i.e., with an average offset of 48 s after starting the evaluation phase and 62 s before entering the verification phase (see Fig. 4).

Our trained model with posterior transition probabilities is shown in Fig. 5. Each state is holding a multivariate GMM which consists of multiple Gaussian mixture distributions (see Fig. 6 left for a univariate example). We showcase the inference of one full exemplary purchase process in Fig. 6 right. Such phase predictions can be

Fig. 3 Experimental VR setup (3D printer decision top, washing powder decision bottom)

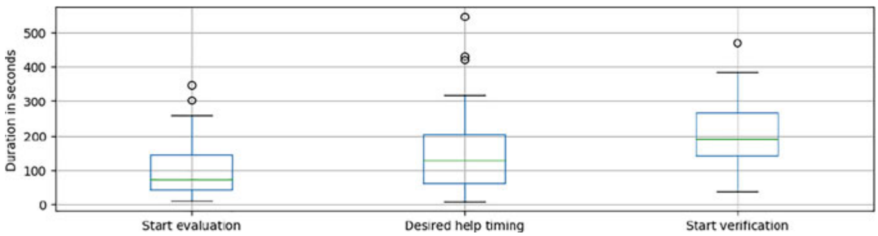


Fig. 4 Boxplots of the self-reported phase shifts and the desired help timing

further refined and leveraged by a UAS or sales agent to find the best time to approach customers with an assistance offering. It is noteworthy that training duration only lasted 3.21 s and with very brief inference times a single observation can be predicted on the fly. The mean difference between classified and reported shifts from orientation to evaluation is −0.14 (SD = 4.49) five second time windows. Overall, the model fits 84.89% of the five second windows correctly.

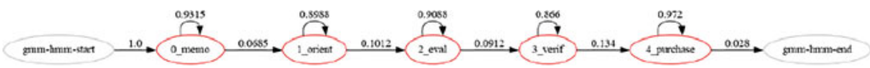


Fig. 5 GMM-HMM with posterior transition probabilities

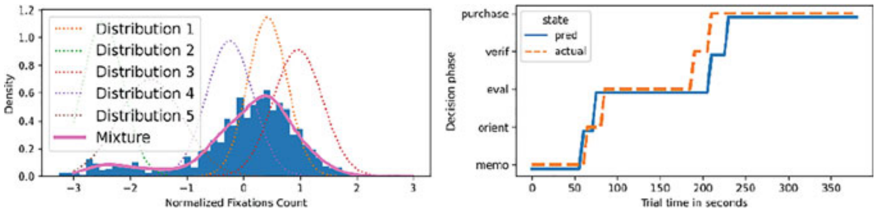


Fig. 6 Exemplified univariate GMM for a single feature (left), comparison between reported state transitions and model prediction for one purchase decision (right)

5 Discussion

Our results show the feasibility of identifying a good timing to approach customers in a virtual commerce scenario using GMM-HMMs and thus yield an answer to our research question. The presented approach uses multiple neurophysiological sensors as input and meets our goal to overcome pure comparison and fixation-based phase determination. The presented methodology can be adopted by other researchers and practitioners to build a maybe soon to be realized overarching virtual platform, offering a multitude of interconnected virtual worlds and services.

This work has limitations which may serve as a guideline for future research. First, our sample almost exclusively consists of students, which limits generalizability. Future research should involve a broader cross-section of society. Second, the sample size should be increased. Our 50 observations yield little variety to equip the model with performant predictive power. Third, immersion, perceived telepresence, and perceived product involvement could have been increased by adding more sensory channels (particularly audio) to the virtual environment. Room size also played a limiting role, as participants had to remain relatively immobile and could not fully immerse themselves in the virtual space. Regarding the applied machine learning techniques, we plan to rigidly quantify the model performance and give detailed information about the most relevant features. We also want to consider further measurements as features, such as electrodermal activity and electroencephalography, which eventually might also be integrated into future HMDs off-the-shelf. Finally, we plan to evaluate the simplified GMM classifier version of the model in an experimental virtual commerce shopping scenario in which a digital human agent approaches a customer according to the timing suggested by the model.

Acknowledgements Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—GRK2739/1—Project Nr. 447089431—Research Training Group: KD2School—Designing Adaptive Systems for Economic Decisions

References

1. Ball, M. (2022). *The Metaverse*. Liveright, [S.I.].
2. Bock, O., Baetge, I., & Nicklisch, A. (2014). Hroot: Hamburg registration and organization online tool. *European Economic Review, 71*, 117–120.
3. Bourlakis, M., Papagiannidis, S., & Li, F. (2009). Retail spatial evolution: Paving the way from traditional to metaverse retailing. *Electronic Commerce Research, 9*, 135–148.
4. Eddy, S. R. (2004). What is a hidden Markov model? *Nature biotechnology, 22*, 1315–1316.
5. Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1968). A model of consumer motivation and behavior. In: *Research in consumer behavior* (pp. 3–20). Holt, Rinehart and Winston, Inc.
6. Friemel, C., Morana, S., Pfeiffer, J., et al. (2018). On the role of users' cognitive-affective states for user assistance invocation. *Information Systems and Neuroscience* (pp. 37–46). Springer.
7. Gidlöf, K., Wallin, A., Dewhurst, R. et al. (2013). Using eye tracking to trace a cognitive process: gaze behaviour during decision making in a natural environment. *JEMR, 6*.
8. Horvitz, E. J., Breese, J. S., & Heckerman, D. et al. (2013). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. arXiv preprint arXiv:1301.7385
9. Krogh, A., Brown, M., Mian, I. S., et al. (1994). Hidden Markov models in computational biology: Applications to protein modeling. *Journal of molecular biology, 235*, 1501–1531.
10. Lanier, J., & Biocca, F. (1992). An insider's view of the future of virtual reality. *Journal of Communication, 42*, 150–172.
11. Liu, H., Gamboa, H., & Schultz, T. (2023). Sensor-based human activity and behavior research: where advanced sensing and recognition technologies meet. *Sensors, 23*.
12. Maedche, A., Morana, S., Schacht, S., et al. (2016). Advanced user assistance systems. *Business & Information Systems Engineering, 58*, 367–370.
13. Makowski, D., Pham, T., Lau, Z. J., et al. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research, 53*, 1689–1696.
14. Patney, A., Salvi, M., Kim, J., et al. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph, 35*, 1–12. https://doi.org/10.1145/2980179.2980246
15. Pedregosa, F., Varoquaux, G., Gramfort, A. et al. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning Research, 12*:, 825–2830.
16. Peukert, C., Lechner, J., Pfeiffer, J., et al. (2020). Intelligent invocation: Towards designing context-aware user assistance systems based on real-time eye tracking data analysis. *Information Systems and Neuroscience* (pp. 73–82). Springer.
17. Pfeiffer, J. (2011). *Interactive decision aids in e-commerce*. Springer Science & Business Media.
18. Pfeiffer, J., Pfeiffer, T., Meißner, M., et al. (2020). Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research, 31*, 675–691.
19. Pham, T., Lau, Z., Chen, S. H., et al. (2021). Heart rate variability in psychology: A review of HRV indices and an analysis tutorial. *Sensors, 21*, 3998.
20. Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine, 3*, 4–16.
21. Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *The Journal of Consumer Research, 21*, 274.
22. Schreiber, J. (2017). Pomegranate: fast and flexible probabilistic modeling in python. *The Journal of machine Learning Research, 18*, 5992–5997.
23. Schultz, T., & Waibel, A. (2001). Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Communication, 35*, 31–51.
24. Sihi, D. (2018). Home sweet virtual home. *JRIM, 12*, 398–417.
25. Stephenson, N. (2003). *Snow crash*. Random House Publishing Group.
26. Sutherland, I. E. (1965). The ultimate display. *Proceedings of the IFIP Congress, 2*, 506–508.
27. Sykes, T. A. (2015). Support structures and their impacts on employee outcomes. *MIS Quarterly, 39*, 473–496.

28. vom Brocke, J., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems, 53*, 1–13.

Can We Replicate Impaired Vision with Simulation Glasses in Computer-Based Task? An Eye Tracking Validation Study

Yasmine Maurice, Félix Giroux, Camille Lasbareilles, Jared Boasen, Sylvain Sénécal, and Pierre-Majorique Léger

Abstract With growing pressure to develop accessible apps and websites, designers or User Experience researchers also face the challenge of recruiting people with disabilities to conduct inclusive usability evaluations. While many researchers rely on disability simulations to identify usability issues, others argue that disability simulations cannot fully replicate the behavior and the lived experience of people with disabilities. This paper presents a study that investigates whether we can replicate the visual search behavior of low vision people on a computer interface with normally sighted participants experiencing a visual disability simulation. A total of 46 participants, including 9 low vision people, and 37 normally sighted participants with and without visual disability simulation glasses, performed computer-based vision tests

Y. Maurice (✉) · F. Giroux · J. Boasen · S. Sénécal · P.-M. Léger
Tech3Lab, HEC Montréal, Montréal, Québec, Canada
e-mail: yasmine.maurice@hec.ca

F. Giroux
e-mail: felix.giroux@hec.ca

J. Boasen
e-mail: jared.boasen@hec.ca

S. Sénécal
e-mail: sylvain.senecal@hec.ca

P.-M. Léger
e-mail: pierre-majorique.leger@hec.ca

P.-M. Léger
Department of Information Technologies, HEC Montréal, Montréal, Québec, Canada

S. Sénécal
Department of Marketing, HEC Montréal, Montréal, Québec, Canada

C. Lasbareilles
Wellcome Centre for Integrative Neuroimaging, Department of Clinical Neuroscience, University of Oxford, Nuffield, England
e-mail: camille.lasbareilles@ndcn.ox.ac.uk

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_20

and visual search tasks. Using eye tracking, we show that the disability simulation tends to replicate the visual search behavior of low vision participants.

Keywords Accessibility · Visual Impairment · Disability Simulation · Eye Tracking · Visual Search

1 Introduction

In 2017, it was reported that over five million people in the U.S. faced challenges accessing information technologies (IT) due to low vision, and this number is expected to nearly double by 2050 due to the aging population [1, 2]. Accordingly, increased efforts have been made to make apps and websites more accessible since the World Wide Web Consortium (W3C) introduced the Web Content Accessibility Guidelines (WCAG) in the mid-1990s. Furthermore, the implementation of laws and policies from governmental agencies and organizations such as the United-Nations have contributed to web accessibility through the implementations of laws [35]. However, as of 2022, 96.8% of the million most important websites' homepages still do not conform to the WCAG 2.0 guidelines [3]. In other words, these websites are still not accessible to people with visual impairments and web inaccessibility remains a pervasive issue [36]. The ramifications of this are important as our reliance on IT continues to grow. Web inaccessibility can create barriers for people with low vision's access to essential activities carried online such as work, education and communication [37].

Consequently, designers are encouraged to include people with disabilities (PWD) in the development cycles of apps or websites. However, they also face the challenge of recruiting visually impaired people to conduct usability evaluations [4, 5]. This population tends to be of older age and suffer from comorbidity that can make the testing of apps and websites increasingly demanding [10, 38]. Furthermore, there is a lot of heterogeneity in the presentations of symptoms which makes recruiting participants with the same level of impairment additionally challenging [8]. Finally, there is increased resource costs associated with recruiting PWD for user testing (e.g., time, money, training) that might not be available to all [39]. A circumvention to this issue is to recruit able-body participants and simulate a disability through the replication of symptoms associated with that disability (e.g., wearing a blindfold to simulate blindness) [39]. This method has been used by scholars in Human Computer Interaction (HCI) and other fields where they attempt to identify usability issues by assessing how able-body participants experiencing a disability simulation interact with a software [6–9]. However, it has been argued that visual disability simulations do not accurately replicate a visually impaired person's behaviors in activities of daily living, including computer-based tasks [10]. These claims are primarily based on low vision participants' self-reports and do not target their use in the context of a participatory user test [10, 11]. Moreover, the goal is not to exactly replicate the experiences of PWD but rather to allow for a first-person perspective through the use of a simulation to get insights about the accessibility of a software with the effects of low vision [40]. Currently, there is a lack of empirical evidence comparing the behaviors of low vision people to those of normally sighted users experiencing a visual disability simulation.

Here, we investigate whether the visual search behavior of low vision people using a computer interface can be replicated with the visual search behavior of normally sighted participants experiencing a visual disability simulation. Due to the impact of low vision on the ability to perceive visual features, we expect to find poorer performance for the low vision participants (simulated and real) compared to controls. In this study, we measure performance through reaction time (RT) and the ratio of fixations over saccades in a visual search task.

Using eye tracking, we assessed the visual search behavior of nine low vision and 38 normally sighted participants in a visual search task. After an initial assessment of the contrast sensitivity (CS) of each group, our results suggests that the visual search behavior of normally sighted people experiencing disability simulation glasses is comparable to the one of people with low vision. This promising finding brings accessibility research one step further towards a better understanding of how visual disability simulations allow normally sighted people to identify usability issues in apps and websites that are experienced by low vision people.

2 Methods

Procedure

This study was conducted in a usability lab in North America between the months of July and December 2022. We chose a between-group design approach to compare our three experimental groups. The experiment procedures began with a 9-point eye tracking calibration. Then, all participants performed the Freiburg vision test (FrACT) CS threshold test in order to compare the extent to which each group can distinguish characters from a background. Subsequently, they performed a visual search task, and followed by a series of naturalistic tasks on a banking website, which will be analyzed in future work. The participants remained at a constant distance of 65 to 70 cm from the screen throughout the experiment. Furthermore, the lighting conditions of the screen and the experimental room were the same for each participant.

Participants

A total of 46 participants aged 21–70 (mean = 41.37; SD = 14.67) were recruited, including 23 males and 23 females. The low vision group included nine participants aged 42–70 (mean = 58; SD = 10.06) with diagnosed low vision conditions. They had a combination of 12 different visual symptoms. All nine had visual acuity (VA) loss and five had CS loss. Then, 37 normally sighted participants were randomly assigned either to the low vision simulation condition or the control condition. There were 25 participants aged 20 to 70 (mean = 37.32; SD = 13.48) wearing the low vision simulation and 12 participants aged 21 to 59 (mean = 37.33; SD = 11.38) in the control condition.

Low Vision Simulation

We adopted the Cambridge simulation glasses [12] to simulate reduced VA and CS. Both measures provide complementary information when evaluating loss of visual function as VA determines the ability to see fine details and CS determines the ability for spatial and pattern discrimination [13]. These glasses have been used in several studies, in which they are typically superimposed to vary the severity of visual impairment [14–18]. Among our 26 normally sighted participants wearing visual disability simulation glasses, 14 wore two pairs of superimposed glasses and 11 wore four pairs [19]. These simulated visual impairments correspond to mild-to-moderate low vision according to the World Health Organization [20]. Two different levels of simulation were randomly assigned to the participants as a way to replicate the heterogeneity of symptoms found in a low vision population.

Contrast Sensitivity Test

All participants' CS threshold was assessed using the standardised, web-based FrACT [21, 22]. The FrACT is a validated test that has been chosen to assess the CS threshold due to its implementability. This test can easily be administered on any computer screen, it is available through Google Chrome and it is open source [21, 22]. CS threshold was assessed using a version of the FrACT computer-based visual test battery, known as the Tumbling E task. This task consists of 24 trials of a single optotype, an image of the letter "E", pointing in four directions (up, down, left, or right, Fig. 1) to determine the "minimum visible" of each participant. Using the keyboard arrows, participants were asked to indicate the direction of the "E", which decreases in contrast (i.e., creating a smaller difference in color between the "E" and the background) following correct answers and increases in contrast again following wrong answers. When participants could not see the direction of the "E", they were instructed to hit an arrow key to the best of their abilities or at random, known as the "forced choice" principle [22]. The test results are provided in the form of a single metric, the LogCS, used by ophthalmologists to quantify CS [23].

Fig. 1 Freiburg vision test (FrACT) answer choices (left), and one example of the contrast sensitivity test image (right)

Visual Search Task

In the present study, we aim to manipulate the contrast of specific colors used on a banking website tested in a subsequent usability test. Object contrast is known to influence reaction time (RT) in visual search tasks [24]. Therefore, to assess RT we developed our own version of a Spatial Configuration Search (or sometimes referred to as a "serial search task") [41]. In the task, participants were exposed to a series of 32 images, each containing 8×8 rows of 64 alphanumeric symbols with one target symbol. In these images, we manipulated the alphanumeric symbol type (i.e., "4" among A's, "2" among Z's, "5" among S's, and "8" among B's, Fig. 2), contrast (i.e., high and low contrast), and target position (i.e., quadrant 1, 2, 3, 4). The high contrast (10.66:1 ratio) was alphanumeric symbols of a dark gray color (#383838 [hexadecimal color]) on a light gray background (#f4f4f4), and the low contrast (1.47:1 ratio) was light gray (#cccbcb) alphanumeric symbols on the same background (#f4f4f4), which represents a contrast ratio that is below the minimum contrast ratio of 4.5:1 for text and images of text following the WCAG 2.1 level AA success criterion 1.4.3 [25]. The 32 image stimuli were presented by a group of alphanumeric symbol types, with the contrast level being randomized. The participants were provided with examples prior to the task. Before each stimulus, participants were instructed to fixate on a cross in the center of the monitor. We used the on-gaze advance function (i.e., 300 ms fixation located in a predefined rectangle shaped area of interest on the target symbol) to measure RT and to move on to the next stimuli. In cases where the eye tracking was not stable, we instructed the participants that the image would change when they find and look at the target symbol, but to say out loud when they found it if the image does not change. The moderator would then manually move on to the next stimulus and the RT would then be based on the time of participants' verbal cue. This task was chosen because no attributes can guide the search for the target. The participant would have to search within the distractors in a random fashion [41]. From this, different visual search patterns could emerge between groups which will be analyzed in future works.

Fig. 2 Visual search task's targets position (left), one example of high contrast image (middle), and one example of low contrast image (right)

Visual Search Behavior

Eye movement measures can objectively inform on one's visual search behavior and use of a website [26–28]. We recorded eye movements at 60fps using a Tobii Pro Nano eye tracking device (Tobii, Karlsrovagen, Sweden). We calibrated the eye tracker using Tobii Pro Lab (Version 1.181) using larger calibration targets with increased contrast. The calibration was successful for all participants except one participant of the low vision group and one participant of the simulation group, which were excluded from the analyses. Eye-tracking was less stable for participants in the simulation group and the low vision group. This is because the glasses' frame of the simulation group and pre-existing conditions such as cataracts of the low vision group impede the ability of the eye tracker to consistently detect the pupil throughout the experiment. Consequently, 11 participants from the simulation group and 2 participants from the low vision group were excluded from the analysis due to data loss across the whole visual search task. For the remaining participants (N = 33), the resulting variance in sampled oculometric data volumes meant that typical visual search behavioral measures such as total fixation and saccade count or total fixation and saccade time [28] could not be used. Therefore, we used the ratio of fixation time to saccade time (i.e., fixation/saccade ratio) over the duration of a visual search task. This ratio allowed us to compare our three groups without being affected by the unequal amount of data lost between the groups during the visual search task. To calculate the ratio, we computed the sum of time of whole fixation and whole saccade captured within each image stimulus' interval. The sum of fixation time was divided by the sum of saccade time in each image, and then aggregated by contrast level for each group. An increasing fixation/saccade ratio reflects more time spent trying to process the local visual information at each fixation (i.e., fixation time) than scanning a visual stimulus (i.e., saccade time). The literature suggests that the fixation/saccade ratio in visual search task increases with poorer image contrast or visibility [24, 29], or in low vision people [30, 31].

Statistical Analysis

We performed five one-way analyses of variance (ANOVA) to assess the differences in the average CS threshold, the average RT with low contrast and with high contrast, and the average fixation/saccade ratio with low contrast and with high contrast, according to our three experimental groups. The one-way ANOVAs on the average CS threshold and average RT were performed with our complete sample, but our one-way ANOVAs on the average fixation/saccade ratio were performed with our participants with available oculometry data. All statistical comparisons were made using SPSS 26 (IBM, Armonk, NY, USA) with significance threshold set at $p \leq 0.05$. Our tests of homogeneity of variance based on mean for our one-way ANOVA were all significant with a threshold set at $p \leq 0.05$, which suggests unequal variances between our group for these variables. Therefore, we ran Welch's robust test of equality of means and adjusted the p-value and F statistics accordingly. Post-hoc comparisons between groups were performed with the Games-Howell procedure.

3 Results

Contrast Sensitivity

Illustrated in Fig. 3, our one-way ANOVA on the CS threshold revealed a significant main effect of group ($Fw(2, 17.6) = 53.355$, $p < 0.001$), with simple main effects testing showing that the control group had significantly higher CS threshold (2.01 ± 0.11 logCS) than both the low vision group (1.35 ± 0.43 logCS; $p = 0.007$) and the simulation group (1.53 ± 0.17 logCS; $p < 0.001$). There was no significant difference between the simulation and the low vision groups ($p = 0.740$). Furthermore, the differences between groups are in line with our expectations, showing that the simulation affects the CS of normally sighted participants to similar levels as the low vision participants.

Reaction Time

For our visual search task (Fig. 3), our one-way ANOVA on the RT in high contrast showed a significant main effect of group ($Fw(2, 17.3) = 20.701$, $p < 0.001$), with simple main effects testing revealing that the control group had significantly faster RT than both the low vision group ($p = 0.002$) and the simulation group ($p < 0.001$). We also found that the simulation group had significantly faster RT than the low vision group ($p = 0.015$) in high contrast stimuli. The one-way ANOVA on the RT in low contrast also showed a significant main effect of group ($Fw(2, 16.1) = 15.227$, $p < 0.001$), with simple main effects testing revealing that the control group had

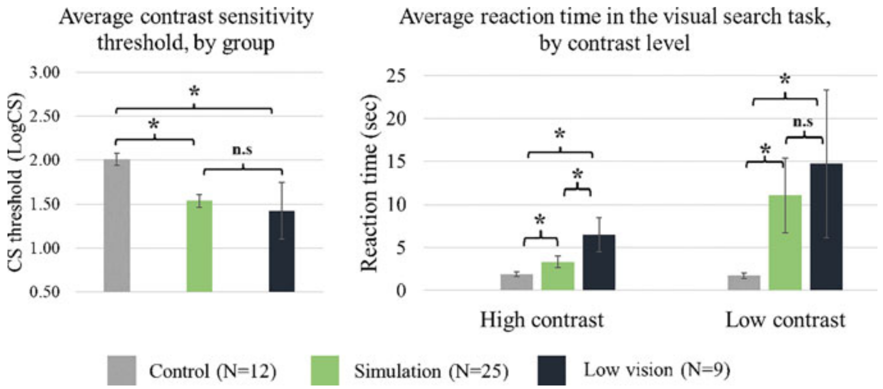


Fig. 3 Average CS (left), average RT for high and low contrast level images in the visual search task (right), by experimental group. * $p < 0.05$

significantly faster RT than both the low vision group ($p = 0.020$) and the simulation group ($p = < 0.001$). However, there was no significant difference between the simulation and the low vision groups ($p = 0.675$) in low contrast stimuli. This shows that the simulation affects RT for both stimuli (high and low contrast), but the effect is bigger for the low contrast stimuli and similar to low vision participants.

Visual Search Behavior

Lastly, for the visual search behavior (Fig. 4), our one-way ANOVA on the fixation/saccade ratio in high contrast images revealed a significant main effect of group ($Fw(2, 10.9) = 8.791$, $p = 0.005$), with simple main effects testing showing that the control group had a significantly lower fixation/saccade ratio than the low vision group ($p = 0.024$), but not the simulation group ($p = 0.078$). The difference between the simulation and low vision groups was also not significant ($p = 0.071$). For low contrast images, our one-way ANOVA on the fixation/saccade ratio also showed a significant main effect of group ($Fw(2, 10.1) = 16.001$, $p < 0.001$), with simple main effects testing showing that the control group had a significantly lower fixation/saccade ratio than both the low vision group ($p = 0.011$) and the simulation group ($p = 0.008$). We found no significant difference between the simulation and low vision groups ($p = 0.445$). This shows that although the simulation had an effect on the number of fixations over the saccades, the effect was stronger for the low contrast stimuli. In the high contrast condition, the visual search behavior of the simulation was not different to the visual search behavior of normally sighted and of low vision participants.

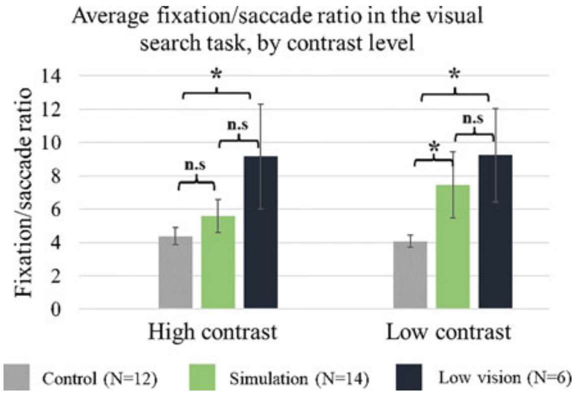


Fig. 4 Average fixation/saccade ratio over the visual search task for high and low contrast level images, by experimental group. * $p < 0.05$

4 Discussion

In this study, we aimed to assess whether the visual search behavior of low vision people using a computer interface can be replicated with normally sighted participants experiencing a visual disability simulation. We show that the simulation glasses were able to replicate, on average, the CS threshold of our low vision participants. Additionally, the CS threshold of the low vision group had a larger deviation from the mean. This can be a result of the symptoms of the participants with low vision where 5 out of 9 reported having lower CS. In the visual search task, our results show that normally sighted participants wearing disability simulation glasses demonstrate similar visual search performance and behavior to those with low vision during interactions with low-contrast stimuli, which is in line with past literature [24, 29–31]. The results indicate that, although the simulation glasses produce poorer vision, the simulation group retained some visual search behaviors while also adopting the behaviors low vision participants during interactions with low-contrast stimuli. The literature suggests that with lower contrast stimuli in visual search the number fixations increases while saccades are not affected [29], and that low vision stimuli are more difficult for low vision people than high contrast ones [30]. Our findings are in line with previous literature on visual search performance in computer tasks and show that the effects of low vision can be replicated through measures of visual search behavior. However, the effect seems so be more important when the condition of the task is more difficult.

This study has three main limitations. First, there was a small number of low vision participants in our sample due to the challenges of recruiting PWD [4, 5]. Second, the data loss resulting from the eye tracker reduced the final sample for visual search task comprised of six low vision participants. Although we were able to find significant differences, the smaller sample may introduce bias in our results. Third, the average age of the low vision participants is 20.68 years older than our normally sighted participants of the control and simulation groups. In fact, it is known that the visual field decreases with age, which may limit the amount of information that can be acquired within one fixation [32].

It should be noted that the LogCS was used to compare our groups and should not be used as a reference for the Cambridge Simulation Glasses as it was not the goal of this study.

Nevertheless, our results suggest that both low vision and simulation groups may have similar visual search behavior in a computer-based information search task. This idea will be investigated in future work where we will analyze the usability issues (e.g., WCAG 2.1 level AA success criterion 1.4.3: Minimum contrast) [25] experienced and reported by our participants in an ongoing naturalistic online banking website study. Our findings will help us better interpret the results yielded from the naturalistic tasks. These new insights will contribute to our understanding on how disability simulations can be used to replicate the behavior of people with disabilities, and consequently the usability issues experienced and identified in user test context [33, 34].

Building on our findings, future research on the use of low vision simulation could investigate the extent to which the effect of the simulation has on the ability to find accessibility issues relating to contrast. Future research questions could expand to naturalistic tasks and look into the differences between simulated low vision and low vision in identifying contrast issues and issuing design recommendations.

References

1. Chan, T., Friedman, D. S., Bradley, C., & Massof, R. (2018). Estimates of incidence and prevalence of visual impairment, low vision, and blindness in the United States. *JAMA ophthalmology, 136*(1), 12–19.
2. Varma, R., Vajaranant, T. S., Burkemper, B., Wu, S., Torres, M., Hsu, C., ... & McKean-Cowdin, R. (2016). Visual impairment and blindness in adults in the United States: demographic and geographic variations from 2015 to 2050. *JAMA ophthalmology*, *134*(7), 802-809.
3. *The WebAIM Million.* (2020). WebAIM. https://webaim.org/projects/million/.
4. Sears, A., & Hanson, V. (2011, May). Representing users in accessibility research. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 2235–2238).
5. Pernice, K., & Nielsen, J. (2001). *How to conduct usability studies for accessibility*. Technical Report, Nielsen Norman Group, USA.
6. Cornish, K., Goodman-Deane, J., & Clarkson, P. J. (2014). Designer requirements for visual capability loss simulator tools: differences between design disciplines. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part I 8* (pp. 19–30). Springer International Publishing.
7. Keates, S., & Looms, P. O. (2014). The role of simulation in designing for universal access. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part I 8* (pp. 54–63). Springer International Publishing.

8. Ates, H. C., Fiannaca, A., & Folmer, E. Immersive simulation of visual impairments using a wearable see-through display. In Proceedings of the ninth international conference on tangible, embedded, and embodied interaction, pp. 225–228. (2015).
9. Tigwell, G. W. (2021, May). Nuanced perspectives toward disability simulations from digital designers, blind, low vision, and color blind people. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
10. Macnamara, A., Chen, C., Schinazi, V. R., Saredakis, D., & Loetscher, T. (2021). Simulating macular degeneration to investigate activities of daily living: A systematic review. *Frontiers in Neuroscience, 15*, 663062.
11. Jones, P. R., & Ometto, G. (2018, March). Degraded reality: using VR/AR to simulate visual impairments. In *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)* (pp. 1–4). IEEE.
12. Clarkson, P. J., Coleman, R., Hosking, I., & Waller, S.D. (2011). Inclusive Design Toolkit, 2nd ed. Available at: www.inclusivedesigntoolkit.com.
13. Xiong, Y. Z., Kwon, M., Bittner, A. K., Virgili, G., Giacomelli, G., & Legge, G. E. (2020). Relationship between acuity and contrast sensitivity: Differences due to eye disease. *Investigative ophthalmology & visual science, 61*(6), 40–40.
14. Goodman-Deane, J., Waller, S., Cornish, K., & Clarkson, P. J. (2014). A simple procedure for using vision impairment simulators to assess the visual clarity of product features. In *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part I 8* (pp. 43–53). Springer International Publishing.
15. Dos Reis, C. S., Soares, F., Bartoli, G., Dastan, K., Dhlamini, Z. S., Hussain, A., ... & Thompson, J. D. (2020). Reduction of visual acuity decreases capacity to evaluate radiographic image quality. *Radiography*, *26*, S79–S87.
16. Stalin, A., & Dalton, K. (2021). Exploration of the minimum visual disability criteria for Para nordic and Para alpine skiing using simulated vision impairments. *Journal of Sports Sciences, 39*(sup1), 167–187.
17. Zallio, M., Waller, S., Chivaran, C., & Clarkson, P. J. (2021). Visual accessibility and inclusion. An exploratory study to understand visual accessibility in the built environment.
18. Angeleska, E., Aleksovska, A., Avramov, N., Sidorenko, S., Rizov, T., & Jankovic, A. (2022). Design and evaluation of an inclusive autonomous vehicle user interface developed for persons with visual acuity loss. *Proceedings of the Design Society, 2*, 2035–2044.
19. Goodman-Deane, J., Waller, S., Collins, A. C., & Clarkson, J. (2013, April). Simulating vision loss: what levels of impairment are actually represented?. In *Contemporary Ergonomics and Human Factors 2013: Proceedings of the international conference on Ergonomics & Human Factors 2013, Cambridge, UK, 15–18 April 2013* (p. 347). CRC Press.
20. World Health Organization. (2019). 9D90 Vision impairment including blindness. https://icd.who.int/browse11/lm/en#/, http://id.who.int/icd/entity/1103667651 (2019)
21. Bach, M. (1996). The freiburg visual acuity test-automatic measurement of visual acuity. *Optometry and Vision Science, 73*(1), 49–53.
22. Bach, M. (2006). The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology, 245*, 965–971.
23. Daiber, H. F., & Gnugnoli, D. M. (2020). Visual acuity.
24. Näsänen, R., Ojanpää, H., & Kojo, I. (2001). Effect of stimulus contrast on performance and eye movements in visual search. *Vision Research, 41*(14), 1817–1824.
25. World Wide Web Consortium. (2018). Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/
26. Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*(6), 631–645.
27. Haesner, M., Chandra, J., Steinert, A., & Steinhagen-Thiessen, E. (2018). An eye movement analysis of web usability: Differences between older adults with and without mild cognitive impairment. *Assistive Technology, 30*(2), 66–73.

28. Cowen, L., Ball, L. J., & Delin, J. (2002). An eye movement analysis of web page usability. *People and Computers*, 317–336.
29. Näsänen, R., & Ojanpää, H. (2003). Effect of image contrast and sharpness on visual search for computer icons. *Displays, 24*(3), 137–144.
30. MacKeben, M., & Fletcher, D. C. (2011). Target search and identification performance in low vision patients. *Investigative Ophthalmology & Visual Science, 52*(10), 7603–7609.
31. Senger, C., Margarido, M. R. R. A., De Moraes, C. G., De Fendi, L. I., Messias, A., & Paula, J. S. (2017). Visual search performance in patients with vision impairment: A systematic review. *Current Eye Research, 42*(11), 1561–1571.
32. Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *JOSA A, 5*(12), 2210–2219.
33. Giroux, F., Boasen, J., Stagg, C. J., Sénécal, S., Coursaris, C., & Léger, P. M. (2021). Motor dysfunction simulation in able-bodied participants for usability evaluation of assistive technology: a research proposal. In *Information Systems and Neuroscience: NeuroIS Retreat 2021* (pp. 30–37). Springer International Publishing.
34. Giroux, F., Couture, L., Lasbareille, C., Boasen, J., Stagg, C. J., Fleming, M. K., ... & Léger, P. M. (2022). Usability evaluation of assistive technology for ICT accessibility: Lessons learned with stroke patients and able-bodied participants experiencing a motor dysfunction simulation. In *Information Systems and Neuroscience: NeuroIS Retreat 2022* (pp. 349-359). Springer International Publishing.
35. Lazar, J. (2019). Web accessibility policy and law. *Web Accessibility: A Foundation for Research*, 247–261.
36. Yan, S., & Ramachandran, P. G. (2019). The current status of accessibility in mobile apps. *ACM Transactions on Accessible Computing (TACCESS), 12*(1), 1–31.
37. Babu, R., Singh, R., & Ganesh, J. (2010). Understanding blind users' Web accessibility and usability problems. *AIS Transactions on Human-Computer Interaction, 2*(3), 73–94.
38. McLean, G., Guthrie, B., Mercer, S. W., & Smith, D. J. (2014). Visual impairment is associated with physical and mental comorbidities in older adults: A cross-sectional study. *BMC Medicine, 12*(1), 1–8.
39. Mack, K., McDonnell, E., Jain, D., Lu Wang, L., E. Froehlich, J., & Findlater, L. (2021, May). What do we mean by "accessibility research"? A literature survey of accessibility papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18).
40. Choo, K. T. W., Balan, R. K., & Lee, Y. (2019, May). Examining augmented virtuality impairment simulation for mobile app accessibility design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–11).
41. Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50*(14), 1304–1311.

Seeing Is Feeling: Emotional Cues in Others' Heart Rate Visualizations

Anke Greif-Winzrieth, Verena Dorner, Fabian Wuest, Paul M. Gorny, and Christof Weinhardt

Abstract As more apps and interfaces use visualizations of other users' biosignals to enhance non-verbal communication, we need to improve our understanding of how users interpret these cues. This study focuses on heart rate visualizations derived from electrocardiogram (ECG) signals. It contributes to the literature by summarizing prior insights on how users interpret two cues in particular: the level of heart rate and the extent of variations in heart rate. We conduct an online experiment with 66 participants based on actual ECG data. We investigate whether users associate certain cue combinations with specific emotions, whether we can observe differences in empathy-related measures, and whether users can form correct inferences on actions and events associated with heart rates. Our preliminary analysis indicates that one previously reported result is also observed in our study: elevated heart rate is associated with negative inferences about mood.

Keywords Biofeedback · Electrocardiogram (ECG) · Heart rate (HR) · Visualization · Experiment

A. Greif-Winzrieth (✉) · F. Wuest · P. M. Gorny · C. Weinhardt
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: anke.greif-winzrieth@kit.edu

F. Wuest
e-mail: fabian.wuest@student.kit.edu

P. M. Gorny
e-mail: paul.gorny@kit.edu

C. Weinhardt
e-mail: weinhardt@kit.edu

V. Dorner
Vienna University of Economics and Business (WU), Vienna, Austria
e-mail: verena.dorner@wu.ac.at

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_21

1 Introduction

As the availability of wearable biosensors increases [1], so does the number of applications outside the medical domain that integrate biofeedback to inform users about their own or other users' physical or mental states and help them change certain behaviors [2]. At the same time, technological progress and the emergence of new measurement devices and methods such as smart clothing, smart speakers, or ear canal measurements create new opportunities for NeuroIS researchers [3], especially for taking studies currently performed in the laboratory to the field [4–6].

Intended advantages of self live biofeedback include helping users regulate their emotions [2, 7], stress management [7], or support exercise [8] and learning activities [9]. Applications of foreign live biofeedback include increasing social presence and social connectedness [10, 11] and facilitating group experiences in online settings [12, 13]. The physiological measures underlying many live biofeedback applications are often derived from changes in heart rate (HR) [2, 7]. From a practical point of view, HR sensors are currently the least expensive, most easily accessible, and user-friendly biosensors available [1]. They are already integrated into many applications [2, 7], and new methods for the acquisition of HR and heart rate variability (HRV) data are emerging [3].

From a conceptual point of view, the choice of HR as a feedback signal often rests on the assumption that people intuitively understand how HR and mental states influence each other [14]. But especially for foreign live biofeedback applications, it is unclear to which degree and under which circumstances that assumption holds [15]. This leads to our research questions:

RQ1: Do people associate certain HR patterns with specific emotions?

RQ2: Can they accurately infer actions and events associated with the visualized HR?

Our research aims to contribute to the NeuroIS research agenda in neuro-adaptive system design [16] and the general IS research agenda by improving our understanding of the physical and digital reality of foreign live biofeedback. We, therefore, aim to generate insights into the representation and interpretation of (psycho-) physical data in the digital world [17].

Specifically, we investigate how users perceive HR-based foreign biofeedback and which inferences they draw from the cues inherent in biosignal visualization. First, we conduct a literature analysis to identify potential cues and discuss potential inferences users may make regarding the underlying emotional state. Second, we design and conduct an online experiment with sixty-six participants to test how users perceive and interpret two of these cues and whether they can predict user actions based on foreign live biofeedback. We use actual ECG data and observations from a previous experiment on financial decision-making since finance is a decision context that has been shown to lead to variations in stress levels, emotions, and heart rate [18].

2 Representation and Interpretation of HR-Based Foreign Live Biofeedback Visualizations

Changes in HR can occur as a reaction to a change in mental and emotional states [19]. In live biofeedback applications, they often serve as a proxy for changes in attention, stress, emotional arousal, and cognitive effort [2, 16].

Research has focused on psychological, biological, and technical aspects, for example, accurately measuring HR, and connecting them to the appropriate mental states [2, 16]. A recent study conceptualizes foreign live biofeedback **as a new communication medium**. It applies insights from communication and media research on the non-verbal cues' role in interpersonal communication to its design [12]. Communication aims of biofeedback design choices include ease of understanding and interpretation as well as fostering interpersonal connectedness and empathy in digital settings [2, 10, 12].

Factors influencing how well people interpret foreign live biofeedback include visualization, situational factors, relationship with the other person, and knowledge and beliefs about the signals (see [12] for an overview). For the present study, we focus on **visualization**.

Visualization is the most frequent type of feedback in live biofeedback applications [2], with elements that represent human physiology (e.g., hearts or silhouettes), elements from nature (e.g., trees), geometric shapes (e.g., bars), colors (e.g., green to red), or numbers. Signal data are displayed with (e.g., healthy tree meaning low levels of stress) or without interpretation (e.g., heart rate numbers only) [2].

Observing another person's **elevated HR** can "transmit[s] negative cues about mood" [20, p. 31], and in adversarial settings, even a normal HR can lead observers to make negative inferences about the other person [20]. In social dilemma games, observing an elevated HR can even reduce cooperative behavior [21]. Compared to a consistent HR, **varied HR** can lead to observers reporting "higher perceived other-arousal, empathy, and social presence" [22, p. 688]. Skeuomorphic representations, (i.e., representations that mimic their real-world counterparts) in virtual reality, specifically a beating heart, help observers better than other representations to identify arousal states accurately. Still, the effectiveness depends on the situation, relationship, and VR space [15].

Empirical NeuroIS research on HR and HRV has predominantly investigated arousal [23], for example in auction games ([24–26], see [23] for an overview). However, [27] investigated self-reported and physiological arousal and valence measures in the context of digital communication. They find that self-report measures for arousal tend to correspond with the respective physiological measures while self-reported valence does not match the corresponding physiological measures.

To the best of our knowledge, no study has experimentally and systematically investigated which meaning observers of other people's HR visualizations attach to specific patterns, except for high versus low HR levels and variance. **Our study aims to replicate previous findings, namely that elevated HR is interpreted as**

higher arousal and a negative cue for mood. We combine this with a comparison between varied and consistent HR to create a baseline for the systematic investigation of the interpretation of other patterns in other contexts.

3 Experiment Design and Preliminary Result

Design

We conduct an online experiment (current experiment; CE). Foreign biofeedback visualizations are created with actual ECG data from a previous experiment (PE) on financial decision-making.[1] The PE was conducted with 211 participants, which is a comparably large sample for HR and HRV NeuroIS studies [3]. Physiological data was collected with PLUX ECG sensors [28] that provide accurate and reliable HR data [29].

PE procedure. The PE included two stages. Participants decided how much they wanted to invest in a risky asset in the decision stage. In the result stage, they saw a simulation of the asset's development and the result of their investment decision (gain or loss). Their payout depended on the result. After each stage, participants indicated their arousal and valence states using the Self-Assessment Manikin (SAM) [30] (arousal: 0 = very calm, 4 = very aroused; valence: 0 = very unhappy, 8 = very happy). The SAM scale is widely used due to its brevity and because it is image-based, thus language-free and easy to understand [31].

Generating the foreign biofeedback from PE data. Physiological arousal is computed as the normalized level of arousal a PE participant experienced regarding their physiological activation. Current HR was divided through the baseline HR, measured at the beginning of the experiment over 5 min, during which participants were instructed to relax ($\Theta HR = HR/HR_{baseline}$). Participants were assigned a level of physiological arousal ranging from 1 (low) to 5 (high) for each second [25]. Data quality was assessed ex-post, and all participants whose heart rate signals contained too much noise were excluded. Changes in the level of arousal are visualized as changes in color (dark blue = low arousal; red = high arousal) [25, 32] and changes in the fill level of a human-shaped avatar (Fig. 1).

This study focuses on two cues: the level of heart rate and the extent of variations in heart rate (as the sum of changes in levels of physiological arousal during the respective stage of the PE). We calculate the quantiles for the measures of both cues and choose PE observations representing each cue combination (Table 1). Table 4 in the Appendix provides more details about the thresholds and values the visualizations are based on.

[1] Unpublished study; submission in preparation. For further details on the experiment, please contact the authors.

Fig. 1 Visualization of foreign biofeedback

Table 1 Cue combinations visualized in the online experiment (CE)

Level of HR → Extent of HR variations ↓	Low (L) (below median)	High (H) (above median)
Low (L) (below median)	LowLow (LL)	HighLow (HL)
High (H) (above median)	LowHigh (LH)	HighHigh (HH)

Procedure of current experiment (CE). First, participants receive information about the decision task in the PE and answer two comprehension questions. Second, they watch the visualization of four PE participants' HR recorded during the PE's decision (Decision Group) or the result stage (Result Group). Cue combinations (Table 1) are presented once in random order (30 s each). Following each cue combination, CE participants indicate with SAM (scale as in PE) how they think the PE participant rated their arousal and valence and how they think the PE participant experienced each of the six basic emotions, i.e., happiness, sadness, disgust, fear, surprise, and anger (1 = not at all, 7 = all the time) [33]. Third, participants in the Decision Group guess how much the PE participant invested in the asset. Participants in the Result Group guess the outcome of the PE participant's investment decision (the asset's gain/loss and final value). Finally, CE participants answer the interpersonal reactivity index (IRI) questionnaire [34] and demographic questions (age, gender, experience with HR monitoring). Their payout is a flat fee of £3.00 and a bonus payment of £0.10 for each correct estimate of PE participants' arousal and valence perceptions.

Results

Sample. Sixty-six participants recruited from Prolific, an online platform to recruit research participants from a pool of over 150,000 registered users [35], participated in the experiment. Table 2 provides an overview of the sample demographics and

Table 2 Sample demographics for the current experiment (N = 66)

Gender	
Female	47%
Male	51.5%
Other	1.5%
Age	
Mean (SD)	27.5 (9.18)
Median	24
Min–max	18–70
Experience with monitoring HR signals	
Yes	34.8%
No	65.2%

shows that our study is comparable to other HR and HRV NeuroIS studies with regard to gender and age [3]. The average experiment duration was 16.4 min (SD = 9.4), with an average payout of £3.08. The experiment was implemented in oTree, an open-source python-based framework for implementing interactive experiments [36]. Thirty-three participants were randomly assigned to the Decision Group and Result Group, respectively. There are no significant differences between the Decision and Result Group regarding age, gender, and previous experience with monitoring HR signals.[2]

Preliminary result. Table 3 shows the average estimates of valence and arousal for high versus low extent of variations in HR and high versus low levels of HR. We find no significant differences in the estimates of arousal and valence for low versus high extent of HR variations. Thus we cannot replicate the findings concerning perceived other-arousal reported by Li et al. [22]. One possible reason is the fact that our study used a different context. For the level of HR, we find that arousal estimates are higher for high HR levels than for low HR levels. Valence estimates are higher for low HR levels than for high HR levels. We thus replicate the findings reported by Merrill and Cheshire [20], namely that elevated heart rate is interpreted as higher arousal and a negative cue for mood.[3]

4 Next Steps and Future Research

Planned data analysis. To investigate the relative importance of cue patterns (HR level and extent of HR variations), context (decision vs. result), and experience in HR monitoring, we will conduct two regressions with robust standard errors, with

[2] All p-values > 0.2 using appropriate non-parametric tests.

[3] There are no differences between Decision and Result Group (all p-values > 0.6 using appropriate non-parametric tests).

Table 3 CE participants' estimate of PE participant's arousal and valence self-reports

Variable	Mean	SD	Mean	SD	Test	PostHoc
Extent of HR variations	*Low (L)*		*High (H)*			
SAM arousal	2.023	1.797	2.189	1.099	F = 0.826	
SAM valence	4.114	2.695	4.432	1.804	F = 1.271	
Level of HR	*Low (L)*		*High (H)*			
SAM arousal	1.045	1.069	3.167	1.02	F = 271.992***	L < H***
SAM valence	4.795	1.873	3.75	2.551	F = 14.403***	H < L***
Statistical significance markers: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$						

arousal and valence as dependent variables and controlling for round and demographic factors. Then, we will investigate whether some cue combinations are harder to interpret by computing and comparing the accuracy of the arousal and valence estimates. Analysis of differences in emotion inferences will follow, as well as whether empathy, as measured in the interpersonal reactivity index questionnaire (IRI) [34], impacts the accuracy of arousal and valence estimates. Finally, we will test how well CE participants could infer—based only on the cues inherent in the HR patterns—PE participants' decisions and investment outcomes, respectively.

Future research. We will continue our literature review to identify admissible additional cue patterns and inferences and how they might vary with context, relationship, and other factors. Second, we aim to identify and develop other visualizations for foreign biofeedback and test them within the framework of our current experiment. In this course we also plan to review and adjust details in the current visualization, for example the colors that can significantly affect the perception of self and foreign live biofeedback visualizations [2]. Finally, we plan to apply our best-performing visualizations in an experiment with participant interaction in a lab-in-the-field approach to test our findings' ecological validity.

Appendix

See Table 4.

Table 4 Levels of HR and extent of HR variations for the cue combinations in the decision and result group. Levels of HR are transformed to a scale ranging from 0 (level 1 in Fig. 1) to 100 (level 5 in Fig. 1) and the extent of variations is calculated based on this scale

	LowLow	LowHigh	HighLow	HighHigh
Decision group				
Mean level of HR (median = 23.33)	7.29	23.33	70.3	45.36
Extent of HR variations (median = 9)	2	15	4	14
Result group				
Mean level of HR (median = 17.43)	8.3	11.38	91.32	39.05
Extent of HR variations (median = 8)	4	11	6	14

References

1. Stangl, F., & Riedl, R. (2022). Measurement of heart rate and heart rate variability with wearable devices: A systematic review. In *Wirtsch 2022 Proceedings*.
2. Lux, E., Adam, M. T. P., Dorner, V., Helming, S., Knierim, M. T., & Weinhardt, C. (2018). Live biofeedback as a user interface design element: A review of the literature. *Communications of the Association for Information Systems,* 257–296. https://doi.org/10.17705/1CAIS.04318
3. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A review of NeuroIS research with a focus on applied methods. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 269–283). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_28
4. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 51,* 13–54. https://doi.org/10.1145/3410977.3410980
5. Greif-Winzrieth, A., Peukert, C., Toreini, P., & Adam, M. T. P. (2021). Exploring the potential of NeuroIS in the wild: Opportunities and challenges of home environments. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 38–46). Springer International Publishing. https://doi.org/10.1007/978-3-030-88900-5_5
6. Greif-Winzrieth, A., Peukert, C., Toreini, P., & Weinhardt, C. (2022). The view of participants on the potential of conducting NeuroIS studies in the wild. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 123–131). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_13
7. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS: Information systems and the brain.*
8. de Oliveira, R., & Oliver, N. (2008). TripleBeat: Enhancing exercise performance with persuasion. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services* (pp. 255–264). ACM. https://doi.org/10.1145/1409240.1409268
9. Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (2015). Exploring feedback strategies to improve public speaking: An interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing—UbiComp '15* (pp. 1143–1154). ACM Press. https://doi.org/10.1145/2750858.2806060

10. Stepanova, E. R., Desnoyers-Stewart, J., Höök, K., & Riecke, B. E. (2022). Strategies for fostering a genuine feeling of connection in technologically mediated systems. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–26). ACM. https://doi.org/10.1145/3491102.3517580
11. Järvelä, S., Kätsyri, J., Ravaja, N., Chanel, G., & Henttonen, P. (2016). Intragroup emotions: Physiological linkage and social presence. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00105
12. Feijt, M. A., Westerink, J. H. D. M., De Kort, Y. A. W., & IJsselsteijn, W. A. (2023). Sharing biosignals: An analysis of the experiential and communication properties of interpersonal psychophysiology. *Human–Computer Interaction, 38,* 49–78. https://doi.org/10.1080/07370024.2021.1913164
13. Liu, F., Esparza, M., Pavlovskaia, M., Kaufman, G., Dabbish, L., & Monroy-Hernández, A. (2019). Animo: Sharing biosignals on a smartwatch for lightweight social connection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3,* 1–19. https://doi.org/10.1145/3314405
14. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*, 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010
15. Lee, S., El Ali, A., Wijntjes, M., & Cesar, P. (2022). Understanding and designing avatar biosignal visualizations for social virtual reality entertainment. In *CHI Conference on Human Factors in Computing Systems* (pp. 1–15). ACM. https://doi.org/10.1145/3491102.3517451
16. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*, 9–24. https://doi.org/10.1080/0960085X.2019.1708218
17. Recker, J., Lukyanenko, R., Jabbari, M., Samuel, B. M., & Castellanos, A. (2021). From representation to mediation: A new agenda for conceptual modeling research in a digital world. *MIS Quarterly, 45,* 269–300. https://doi.org/10.25300/MISQ/2021/16027
18. Fenton-O'Creevy, M., Lins, J. T., Vohra, S., Richards, D. W., Davies, G., & Schaaff, K. (2012). Emotion regulation and trader expertise: Heart rate variability on the trading floor. *Journal of Neuroscience, Psychology, and Economics, 5,* 227–237. https://doi.org/10.1037/a0030364
19. Cacioppo, J., Tassinary, L., & Berntson, G. (2007). *Handbook of psychophysiology*. https://doi.org/10.13140/2.1.2871.1369
20. Merrill, N., & Cheshire, C. (2016). Habits of the heart (rate): Social interpretation of biosignals in two interaction contexts. In *Proceedings of the 19th International Conference on Supporting Group Work* (pp. 31–38). ACM. https://doi.org/10.1145/2957276.2957313
21. Merrill, N., & Cheshire, C. (2017). Trust your heart: Assessing cooperation and trust with biosignals in computer-mediated interactions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 2–12). ACM. https://doi.org/10.1145/2998181.2998286
22. Li, B. J., Bailenson, J. N., Ogle, E., & Zaki, J. (2021). Exploring the heart rate as a chronemic cue in virtual settings: How perceptions of consistent and varied heart rates of a storyteller influence self-reported other-arousal, empathy and social presence. *Media Psychology, 24*, 688–712. https://doi.org/10.1080/15213269.2020.1788394
23. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability in NeuroIS research: Review of empirical results. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 285–299). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_29
24. Adam, M. T. P., Krämer, J., & Weinhardt, C. (2012). Excitement up! Price down! Measuring emotions in Dutch auctions. *International Journal of Electronic Commerce, 17*, 7–40. https://doi.org/10.2753/JEC1086-4415170201
25. Astor, P. J., Adam, M. T. P., Jerčić, P., Schaaff, K., & Weinhardt, C. (2013). Integrating biosignals into information systems: A NeuroIS tool for improving emotion regulation. *Journal of Management Information Systems, 30*, 247–278. https://doi.org/10.2753/MIS0742-1222300309

26. Teubner, T., Adam, M., & Riordan, R. (2015). The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions. *Journal of the Association, 16,* 838–879. https://doi.org/10.17705/1jais.00412
27. Walla, P., & Lozovic, S. (2020). The effect of technology on human social perception: A multimethods NeuroIS pilot investigation. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. Randolph, & T. Fischer (Eds.), *Information systems and neuroscience* (pp. 63–71). Springer International Publishing. https://doi.org/10.1007/978-3-030-28144-1_7
28. *Electrocardiography (ECG) sensor*. Retrieved April 09, 2023, from https://www.pluxbiosignals.com/products/electrocardiography-ecg-sensor-1
29. Wilson, F., McHugh, C., MacManus, C., Baggish, A., Tanayan, C., Reddy, S., Wasfy, M. M., & Reilly, R. B. (2022). Diagnostic accuracy of a portable ECG device in rowing athletes. *Diagnostics, 12*, 2271. https://doi.org/10.3390/diagnostics12102271
30. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*, 49–59. https://doi.org/10.1016/0005-7916(94)90063-9
31. Bynion, T.-M., & Feldner, M. T. (2020). Self-assessment manikin. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 4654–4656). Springer International Publishing. https://doi.org/10.1007/978-3-319-24612-3_77
32. Lux, E. (2017). *Live biofeedback in electronic markets.*
33. Ekman, P. (1999). Basic emotions. *Handbook of Cognition and Emotion, 98*, 16.
34. Keaton, S. A. (2017). Interpersonal Reactivity Index (IRI): (Davis, 1980). In D. L. Worthington & G. D. Bodie (Eds.), *The sourcebook of listening research* (pp. 340–347). Wiley. https://doi.org/10.1002/9781119102991.ch34
35. *Prolific · Quickly find research participants you can trust*. Retrieved April 09, 2023, from https://www.prolific.co/
36. Chen, D. L., Schonger, M., & Wickens, C. (2016). OTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

How Experts Rely on Intuition in Medical Image Annotation—A Study Proposal

Florian Leiser, Simon Warsinsky, Manuel Schmidt-Kraepelin, Scott Thiebes, and Ali Sunyaev

Abstract Contemporary machine learning (ML) research discusses the benefits of including domain knowledge in data-driven models under the term informed ML. While scientific domain knowledge can be formalized and integrated easily, expert knowledge is rather tacit and informal. Intuition is considered a key driver of expert judgment but is especially difficult to measure and formalize. In this study, we propose a cognitive task analysis-inspired approach to investigate the role of intuition during medical image annotation with the aid of neurophysiological measurements. We aim to observe 15 experts during their annotation and analyze EEG and eye-tracking data to identify cues indicating intuition. This study should provide insights into expert decision-making and the role of intuition therein and serve as a first step toward a later formalization of expert judgment for expert-informed ML models.

Keywords Intuition · Expert decision-making · EEG · Eye-tracking · Informed machine learning · Medicine

F. Leiser · S. Warsinsky · M. Schmidt-Kraepelin · S. Thiebes · A. Sunyaev (✉)
Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: sunyaev@kit.edu

F. Leiser
e-mail: florian.leiser@kit.edu

S. Warsinsky
e-mail: simon.warsinsky@kit.edu

M. Schmidt-Kraepelin
e-mail: manuel.schmidt-kraepelin@kit.edu

S. Thiebes
e-mail: scott.thiebes@kit.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_22

1 Introduction

We are currently experiencing a rapid rise of machine learning (ML) models. While researchers and practitioners generally recognize the massive potential of ML models in domains like healthcare or finance, questions regarding their accuracy, explainability, and transparency remain [1, 2]. Informed ML is a nascent concept in ML research that aims at improving ML models by integrating available domain knowledge into ML pipelines [3]. Informed ML models may for example integrate formalized scientific knowledge (e.g., physical laws [4]) into the model's loss function to reduce training time. One source of knowledge that has been particularly interesting for scholars is expert knowledge [5], that is, informal knowledge held by a particular group of domain experts [3]. By integrating expert knowledge into ML models, research has reported benefits like more explainable predictions [6] or less data being required for predictions [7]. However, expert knowledge is often difficult to acquire and is usually quite informal (i.e., it must be formalized to be integrated into ML models) [3], therefore, formalizing expert knowledge can be difficult. One reason for this is that the decision-making processes and judgments of experts are often driven by intuition [8, 9]. Intuition is considered a decision-making behavior where individuals are unable to describe the reasoning or other processes that produced the answer in detail [10]. Expertise is one of the main causes of intuition [11] and some scholars argue that intuition sets apart expert and non-expert judgments [9, 12]. However, with the current state of knowledge, we are unable to reliably formalize expert intuition so that it may be integrated into informed ML models.

With this study, we seek to take a first step in formalizing expert intuition to ultimately make it usable for informed ML models. Since intuition is considered context-specific [13] and a key driver of expert decision-making, we select a medical context for our study. More precisely, we turn our attention to the context of expert medical image annotation (MIA). Due to the difficulty of meaningfully interpreting medical images, MIA tasks are usually done by domain experts [14] which have been shown to be potentially erroneous [15]. While there have been first hints on the presence of expert intuition in MIA tasks (e.g., the importance of a "gut feeling" [16, 17]), it remains largely unclear what role intuition plays in how expert medical image annotators make inferences and judgments. Accordingly, we ask: *How do experts rely on intuition in medical image annotation?*

To answer this research question we plan to build on research that has been undertaken to better understand expert intuition [18]. To overcome the varying results of qualitative studies, the domain of neuro-information systems (NeuroIS) recognized great potential to identify intuition using neurophysiological tools like electroencephalography (EEG) or eye-tracking. For example, past NeuroIS research identified prevalent brain areas during intuitive thinking [19] or detected subtle decision-making hints by analyzing eye fixations and frame areas with eye-tracking devices right before the final decision is made [20]. Eye-tracking data is easy to formalize (e.g., by creating attention maps) and, therefore, lends to be integrated into ML models [21].

We propose a research approach where we gather, analyze, and triangulate neurophysiological, interview, self-report, and questionnaire data to infer the role of intuition during these tasks. We contribute to an improved understanding of expert decision-making, especially in the medical context, by validating recent efforts with neurophysiological data. Additionally, we will provide first steps to understand the role of intuition in expert decision-making which are required to formalize intuition in following works.

2 Theoretical Background on Expert Intuition

Intuition describes something subconscious, quick and including no rational, analytical process [22]. Resulting judgments can be "remarkably accurate and sometimes off the mark" [9]. The availability of intuition to an individual is dependent on the level of expertise [23]: Novices may use intuitive thinking as a precursor for analytical understanding (i.e., immature intuition) while experts may rely on "intuitive *seeing* that is accessible once a person has advanced knowledge structures" (i.e., mature intuition) [12]. Therefore, expert intuition differs from non-expert intuition.

One theory that can help to further understand intuition is the dual process theory, which proposes a decision-making process consisting of two systems [8]. An *intuitive* system (System 1), that provides automatic and fast decisions, and a *rational* system (System 2) that performs slower and occupies a higher working memory [24]. With intuition as our research focus, we align ourselves with this perspective and subsequently focus on what is commonly termed by research as Type 1 or System 1 decisions [8, 24, 25]. Most studies on the use of System 1 investigate behavioral aspects [26], but new approaches rely on neurophysiological data, especially in the form of brain imaging tools like functional magnetic resource imaging and EEG, to monitor different characteristics of intuition. These studies found, for instance, increased activity in the insula and amygdala if participants relied on intuition [27] or increased parietal alpha activity when participants faced mathematical tasks [28].

3 Methods

Our research approach is informed by cognitive task analysis (CTA), a family of methods used to study and describe reasoning [29]. CTA consists of three phases namely (1) knowledge elicitation, (2) data analysis, and (3) knowledge representation. An overview of our research design is given in Fig. 1. For our EEG setup, we orient ourselves on existing approaches investigating the role of System 1 in a mathematical task [28] and describe and argue where and why we deviate from that study setup. We combine guidelines on how to report EEG studies in NeuroIS Research [30] with our research design in this section.

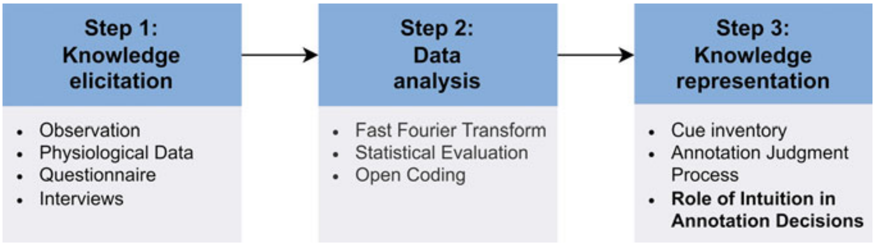


Fig. 1 Research design based on CTA [29]

Knowledge Elicitation

Participants. During the knowledge elicitation we seek to observe experts conducting an MIA task. We aim to include 15 MIA experts in our studies. This number slightly exceeds current guidelines on phenomenological-driven research [31] allowing for errors during data gathering like misfunctioning sensors, abortion of experiments or insufficient performance levels. Phenomenological research investigates human experience, in our case expert intuition [32]. To gauge our participants' expertise, we measure the quality of the annotations they provide by using boundary intersection over union (BIoU) measure of the segmentation masks [33]. To ensure validity, we intend to exclude those participants, where the BIoU score is on average below 0.8 to ensure valid expertise. We suspect our participants to be annotation experts that are generally healthy, with normal or corrected-to-normal vision, no neurological impairments and to be between 20 and 50 years of age with mixed gender. For our purposes, we consider an MIA expert to be someone who is familiar with the current tools, tasks, and standards. Therefore, even young annotators can be seen as experts.

All participants are asked to provide informed consent and earn a 50 € allowance and up to 30 € for their travel expenses. The study was approved by the Ethics Committee of the Karlsruhe Institute of Technology in June, 2023.

Procedure and Instruments. The participants will be greeted in individual sessions where we first give them an overview of the experiment. After all administrative questions of the participants are resolved, we will set up the collection of physiological data in the form of EEG and eye-tracking. To measure eye movements during the annotation we will use the Tobii Eye Tracker 5 (Tobii AB, Stockholm, Sweden). The eye-tracking device needs to be calibrated at first by following the manufacturer's guidelines. EEG data are recorded with a standard gel-based Emotiv Flex (Emotiv Pty Ltd, San Francisco, CA, USA) with 32 Ag/AgCl electrodes evenly spread across using the international 10–20 system. We include one reference electrode (CMS) on the left earlobe and the ground electrode (DRL) on the right earlobe. Electrode impedances are kept below 20 kΩ. Data are sampled at 500 Hz and filtered with the built-in 5th-order sinc filter. Before we start the annotation task, we ask our participants to solve the "add-zero" and "add-one" tasks conducted by Williams et al. [28].

This serves two purposes: It allows our participants to become comfortable with the devices attached to them and by reproducing the existing findings of Williams et al. [28], we can ensure the validity of our setup and measurements.

For the actual annotation task, the participants will segment medical instruments on 20 frames of the publicly available Heidelberg Colorectal data set [34] using an annotation web application [35]. We chose this dataset, as it has ground truth annotations to use as an objective comparison point. The dataset also provides difficulty assessments for each image and is publicly available. During the annotation task, we will observe our participants, record mouse and keyboard inputs, and collect EEG and eye-tracking data. Since we investigate an annotation task, our experts need to move their hands and eyes. We will instruct our participants to move their head as little as possible to avoid movement artifacts during the EEG data gathering as much as possible. We aim to achieve synchronization between the different data collection methods by logging markers after changing frames in the annotation tool and using assisting tools (e.g., Lab Streaming Layer, https://github.com/sccn/labstreaminglayer).

After finishing the annotation task, we present participants with a quick survey including questions on their sociodemographic data (e.g., age, gender, handedness), their experiences during the annotation task (e.g., subjective workload assessed by the NASA Task Load Index [36]) and their general stance toward intuition based on the rational-experiential inventory [37]. Finally, we conduct semi-structured interviews to further delve into possible occurrences of intuition during the annotation tasks. We will design the questions in each individual interview in a way that they target characteristics of intuition that occurred during the annotation task. For example, we may ask participants to justify their annotation decision on a specific image, if we noticed a rapid decision (a key hallmark of intuition) or cases where participants break away from standard procedures. In summary, we expect the whole procedure to last about 60 min.

Data Processing and Analysis

All types of data will be investigated independently at first and then triangulated later. First, we will compare the recorded annotations to the available ground truth annotations by assessing the BIoU score for every expert. Investigating the annotation time required for each frame might provide us with additional insights about the difficulty of frames or the familiarity with the task of our participants.

To process EEG data, we follow recent approaches to detect intuition with EEG [28] and exclude excessively noisy and faulty electrodes automatically first before we down-sample the remaining data to 250 Hz. We first aim to conduct a restricted info-max independent component analysis before we follow with a Fast Fourier transformation without tapering and normalize the output. This allows us to explore the broad range of theta and alpha bands which have been shown to play a crucial role in intuitive or cognitive processes [28]. A frequent artifact arising in EEG data extraction are electrooculogram (EOG) artifacts from eye-blinking activity [38]. We will account for EOG artifacts by placing two reference electrodes which do not measure EEG activity. Additionally, we can detect eye blinks with our eye-tracking measures and follow existing approaches to remove the artifacts from our data [38]. To analyze the eye-tracking data, we transmit the data to Python code which creates attention maps that reveal areas where participants look most during the annotation task [39]. Questionnaire data will be analyzed using basic statistical measures. The recorded audio data of the interviews will be transcribed and analyzed via open coding [40].

Ultimately, we seek to triangulate data from all these different sources to identify intuitive thinking. To do so, we will be particularly on the lookout for characteristics of intuitive thinking in the physiological data (e.g., speedy decisions, regions of interest investigated first) and then try to confirm our suspicions with interview data (e.g., an overwhelming sense of certainty, a particularly difficult-to-annotate image). We then finally assess the general stance of our experts toward intuition with the rational-experiential inventory and evaluate the results to our findings. As intuitive thinking is usually rapid, we think that comparing the physiological data during the first two seconds to the remainder of the annotation time for each frame could be particularly insightful. Doing so may allow us to extract differences in the first impression of each frame compared to later, more cognitive analysis.

Knowledge Representation

For the final step of CTA, we present our findings with aiding visualization. Overall, we aim to develop a decision-making process of experts and indicate the relevance of intuition across the steps. We especially rely on the triangulation of physiological data with qualitative interview data to detect potential contradictions between the perceived choices and the measured physiological data.

4 Discussion and Conclusion

Our study aims to provide novel insights into expert judgments in medical image annotation and the role of intuition therein. We aim to identify cues relevant to intuitive decision-making and their prevalent brain areas during experts' decision-making. These cues could inform ML applications after the formalization of the expert decision-making processes [41]. Acknowledging the relevance of intuition in expert decision-making would also allow for improving working environments [42]. This study also investigates the triangulation of data and the use of neurophysiological data in CTA approaches. Therefore, this study could serve as a baseline for combining neurophysiological and CTA techniques.

We do acknowledge some limitations in our proposed study. Due to the context-specific nature of intuition our study requires a very specialized use case and, therefore, results in a small sample size. We suspect that 15 participants with 20 trials per person is sufficient to identify indications of intuition, but it might not yield detailed results. We aim to improve our approach by investigating other use cases in the future. To investigate our participants in their usual behaviors, we also refrain from manipulating the annotation environment (i.e., manipulating experiments) to tease out intuition. Showing frames for only two seconds could, for example, provide additional insights into intuition. However, this could also reduce the impact of physiological data since eye-tracking attention maps have only limited data. Ultimately, with the data we propose to collect in this study, we seek to build a foundation for the inclusion of expert knowledge into medical informed ML models.

References

1. Wang, F., & Preininger, A. (2019). AI in health: State of the art, challenges, and future directions. *Yearbook of Medical Informatics, 28*(1), 016–026. https://doi.org/10.1055/s-0039-1677908
2. Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy Artificial Intelligence. *Electronic Markets, 31*(2), 447–464. https://doi.org/10.1007/s12525-020-00441-4
3. von Rueden, L., et al. (2023). Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Transactions on Knowledge and Data Engineering, 35*(1), 614–633. https://doi.org/10.1109/TKDE.2021.3079836
4. Karpatne, A., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering, 29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168
5. Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis, 69*, 101985. https://doi.org/10.1016/j.media.2021.101985
6. Radovanovic, S., Vukicevic, M., Kovacevic, A., Stiglic, G., & Obradovic, Z. (2015). Domain knowledge based hierarchical feature selection for 30-day hospital readmission prediction. In J. H. Holmes, R. Bellazzi, L. Sacchi, & N. Peek (Eds.), *Artificial Intelligence in Medicine*. Lecture Notes in Computer Science (Vol. 9105, pp. 96–100). Springer International Publishing. https://doi.org/10.1007/978-3-319-19551-3_11
7. Zhang, Y., et al. (2020). COVID-DA: Deep domain adaptation from typical pneumonia to COVID-19. https://doi.org/10.48550/ARXIV.2005.01577
8. Kahneman, D. (2011). *Thinking, fast and slow* (1st ed.). Farrar, Straus and Giroux.
9. Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755
10. Simon, H. A. (1992). What is an 'explanation' of behavior? *Psychological Science, 3*(3), 150–161. https://doi.org/10.1111/j.1467-9280.1992.tb00017.x
11. Dane, E., & Pratt, M. G. (2009). Conceptualizing and measuring intuition: A review of recent trends. In G. P. Hodgkinson & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology* (1st ed., pp. 1–40). Wiley. https://doi.org/10.1002/9780470745267.ch1
12. Baylor, A. L. (2001). A U-shaped model for the development of intuition by level of expertise. *New Ideas in Psychology, 19*(3), 237–244. https://doi.org/10.1016/S0732-118X(01)00005-8
13. Alexander, J., Mallon, R., & Weinberg, J. M. (2014). Accentuate the negative. In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 31–50). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199927418.003.0002

14. Ward, T. M., Fer, D. M., Ban, Y., Rosman, G., Meireles, O. R., & Hashimoto, D. A. (2021). Challenges in surgical video annotation. *Computer Assisted Surgery, 26*(1), 58–68. https://doi.org/10.1080/24699322.2021.1937320
15. Rädsch, T., Eckhardt, S., Leiser, F., Pandl, K. D., Thiebes, S., & Sunyaev, A. (2021). What your radiologist might be missing: Using machine learning to identify mislabeled instances of x-ray images. In *Proceedings of the 54th Hawaii International Conference on System Sciences, Virtual Conference,* January 2021. https://doi.org/10.24251/HICSS.2021.157
16. Freeman, B., et al. (2021). Iterative quality control strategies for expert medical image labeling. In *Proceedings of the Ninth AAAI Conference on Human Computation and Crowdsourcing, Virtual only*, November 2021 (pp. 60–71).
17. Leiser, F., et al. (2023) Understanding the role of expert intuition in medical image annotation: A cognitive task analysis approach. In *Proceedings of the 56th Hawaiian International Conference on System Sciences, Maui, HI, USA,* January 2023. https://doi.org/10.5445/IR/1000153826
18. Akinci, C., & Sadler-Smith, E. (2012). Intuition in management research: A historical review: Intuition in management research. *International Journal of Management Reviews, 14*(1), 104–122. https://doi.org/10.1111/j.1468-2370.2011.00313.x
19. Lieberman, M. D. (2007). Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology, 58*(1), 259–289. https://doi.org/10.1146/annurev.psych.58.110405.085654
20. Tomanek, K., Hahn, U., Lohmann, S., & Ziegler, J. (2010). A cognitive cost model of annotations based on eye-tracking data. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden,* June 2010 (pp. 1158–1167).
21. Shojaeizadeh, M., Djamasbi, S., Paffenroth, R. C., & Trapp, A. C. (2019). Detecting task demand via an eye tracking machine learning system. *Decision Support Systems, 116*, 91–101. https://doi.org/10.1016/j.dss.2018.10.012
22. Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. *Psychological Inquiry, 21*(4), 279–294. https://doi.org/10.1080/1047840X.2010.517737
23. Raufaste, E., Eyrolle, H., & Mariné, C. (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science, 22*(4), 517–546. https://doi.org/10.1207/s15516709cog2204_4
24. Evans, J. S. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21*(4), 313–326. https://doi.org/10.1080/1047840X.2010.521057
25. Thompson, V. A. (2014). What intuitions are… and are not. In *Psychology of Learning and Motivation* (pp. 35–75). Elsevier. https://doi.org/10.1016/B978-0-12-800090-8.00002-0
26. Liang, T.-P., Chou, Y.-C., & Liu, C.-H. (2020). Neural correlates of dual decision processes: A network-based meta-analysis. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. Randolph, & T. Fischer (Eds.), *Information Systems and Neuroscience*. Lecture notes in information systems and organisation (Vol. 32, pp. 203–210). Springer International Publishing. https://doi.org/10.1007/978-3-030-28144-1_22
27. Kuo, W.-J., Sjöström, T., Chen, Y.-P., Wang, Y.-H., & Huang, C.-Y. (2009). Intuition and deliberation: Two systems for strategizing in the brain. *Science, 324*(5926), 519–522. https://doi.org/10.1126/science.1165598
28. Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage, 189*, 574–580. https://doi.org/10.1016/j.neuroimage.2019.01.048
29. Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press.
30. Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems*, *37*. https://doi.org/10.17705/1CAIS.03746

31. Daniel, J. (2012). *Sampling essentials: Practical guidelines for making sampling choices.* Sage Publications, Inc. https://doi.org/10.4135/9781452272047
32. Wilson, A. (2015). A guide to phenomenological research. *Nursing Standard, 29*(34), 38–43. https://doi.org/10.7748/ns.29.34.38.e8821
33. Reinke, A., et al. (2021). Common limitations of image processing metrics: A picture story. https://doi.org/10.48550/ARXIV.2104.05642
34. Maier-Hein, L., et al. (2021). Heidelberg colorectal data set for surgical data science in the sensor operating room. *Scientific Data, 8*(1), 101. https://doi.org/10.1038/s41597-021-00882-2
35. CVAT.ai Corporation. (2022). Computer vision annotation tool (CVAT). https://github.com/opencv/cvat
36. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9
37. Norris, P., Pacini, R., & Epstein, S. (1998). *The rational-experiential inventory, short form.* Unpublished inventory. University of Massachusetts at Amherst.
38. Maddirala, A. K., & Veluvolu, K. C. (2021). Eye-blink artifact removal from single channel EEG with k-means and SSA. *Science and Reports, 11*(1), 11043. https://doi.org/10.1038/s41598-021-90437-7
39. Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research, 4*(1), 7. https://doi.org/10.1186/s41235-019-0159-2
40. Myers, M. D. (2020). *Qualitative research in business and management* (3rd edn.). Sage.
41. Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research, 32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980
42. Warsinsky, S., Schmidt-Kraepelin, M., Thiebes, S., Wagner, M., & Sunyaev, A. (2022). Gamified expert annotation systems: meta-requirements and tentative design. In A. Drechsler, A. Gerber, & A. Hevner, (Eds.), *The transdisciplinary reach of design science research.* Lecture Notes in Computer Science (Vol. 13229, pp. 154–166). Springer International Publishing. https://doi.org/10.1007/978-3-031-06516-3_12

The Effect of Feedback on Electrophysiological Signal Complexity as a Function of Attachment Style

Dor Mizrahi, Ilan Laufer, and Inon Zuckerman

Abstract Attachment theory has been applied to various domains, including developmental, clinical, and social psychology. It has been instrumental in understanding the mechanisms underlying interpersonal relationships, mental health, and well-being. The attachment profiles can be divided into several styles, but the most basic set comprises two basic attachment styles, *secure* and *insecure*. Today, the current practice of measuring attachment typically involves using self-report questionnaires or interviews. However, the self-report data may be influenced by social desirability or other factors that may conceal or distort the respondents true feelings or opinions. Therefore, in this study, we will try to rely on an objective assessment of the attachment style by analyzing scalp EEG brain recordings. Specifically, we sought to investigate whether signal complexity, derived by using the method of Lempel Ziv Complexity (LZC), could differentiate between insecure and secure attachment styles based on a success or failure feedback given in the context of a flanker task. A significant interaction between attachment style and feedback type was found due to the change in complexity level between success and failure as a function of attachment type. Secure players were associated with an increase in complexity level between success and failure, whereas for insecure players no change was observed between these conditions. These results may be explained by different mechanisms of emotional regulation that are employed by secure and insecure participants. Possibilities for future research were also discussed.

Keywords Attachment theory · EEG · NeuroIS · Data analysis · Lempel–Ziv complexity

D. Mizrahi (✉) · I. Laufer · I. Zuckerman
Department of Industrial Engineering and Management, Ariel University, Ariel, Israel
e-mail: dor.mizrahi1@msmail.ariel.ac.il

I. Laufer
e-mail: ilanl@ariel.ac.il

I. Zuckerman
e-mail: inonzu@ariel.ac.il

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_23

1 Introduction

Attachment theory confirms the importance of relationships and their impact on personal development [1–4]. The attachment styles could be divided into two basic categories, secure and insecure [4]. Individuals with a secure attachment style typically have a favorable view of themselves and their relationships and are comfortable seeking and supporting others. In contrast, those with an insecure attachment style (anxious or avoidant) may experience difficulties forming and maintaining close relationships. Today, the current practice of measuring attachment typically involves using self-report questionnaires or interviews [5–7]. However, the self-report data may be influenced by social desirability or other factors that may conceal or distort the respondents true feelings or opinions.

Studies (e.g. [8, 9]) have shown that EEG signal complexity is a measure of the complexity and variability of neural activity, is sensitive to changes in cognitive processes and emotional states. For instance, individuals with insecure attachment styles tend to exhibit reduced neural complexity in response to stressors compared to those with secure attachment styles. Therefore, it is plausible that attachment style and feedback type may interact to influence EEG signal complexity, potentially shedding light on the neural mechanisms underlying individual differences in cognitive and emotional processing. Further investigation into these relationships may provide important insights into the neural basis of attachment and feedback processes, with potential implications for the development of interventions aimed at improving social and emotional well-being. Therefore, in this study, we will try to rely on an objective assessment of the attachment style by analyzing scalp EEG brain recordings. Specifically, we sought to investigate whether signal complexity, derived by using the method of Lempel Ziv Complexity (LZC), could differentiate between insecure and secure attachment styles based on a success or failure feedback given in the context of a simple cognitive task. LZC provides a measure of the information content and irregularity of EEG signals, and is suitable to characterize the development of activity in high dimensional systems. This can be helpful in characterizing changes in brain activity associated with various neurological conditions or states [10, 11]. Moreover, LZC was used in previous studies for emotion recognition [12, 13]. Additionally, we studied the effect of feedback since it was previously shown that emotion regulation is differentially modulated by attachment style [14].

2 Experimental Design

The study comprised two stages. In the first stage, 96 participants filled out the ECR-R questionnaires [15], composed of 36 items. In the second stage, based on the analysis of the ECR-R data, 27 players out of the 96 were invited to the EEG lab to conduct the second part of the experiment in such a way as to create an equal sample from all the different attachment style groups. In this stage, players were engaged in performing 60 trials of the Eriksen flanker task [16] while the EEG was recorded from their scalp]. The experimental protocol was approved by the institution's IRB committee. All participants signed an informed consent form to participate in the study.

Stage 1—Assessment of Secure and Insecure Attachment Styles Based on the ECR-R

In the first stage, 96 students filled out the ECR-R questionnaire which has two dimensions: anxiety and avoidance. Figure 1a shows the distribution of the ECR-R values. The x-axis denotes the avoidance dimension while the y-axis the anxiety dimension. The values of each of the scales range from 1 to 7. Participants who obtain low values in both dimensions of the ECR-R are considered to have a secure attachment style, whereas those participants who obtain high values on both scales are fearful avoidants (or disorganized). Following the results of the distribution of the ECR-R questionnaire, we have clustered the data into the four main attachment styles using the k-means clustering algorithm [17] on the basis of the elbow method [18]. The analysis of the results (Fig. 1a, b) revealed that the optimal number of clusters is k = 4, as can be seen in Fig. 1b. This number of clusters is in accordance with the four different attachment classifications reported in the scientific literature (e.g. [3]).

In an overview of the clusters appearing in Fig. 1b, it can be seen that the cluster of the green dots represents players characterized by secured attachment, while all other players (denoted by blue, red and purple dots) are considered to be insecure.

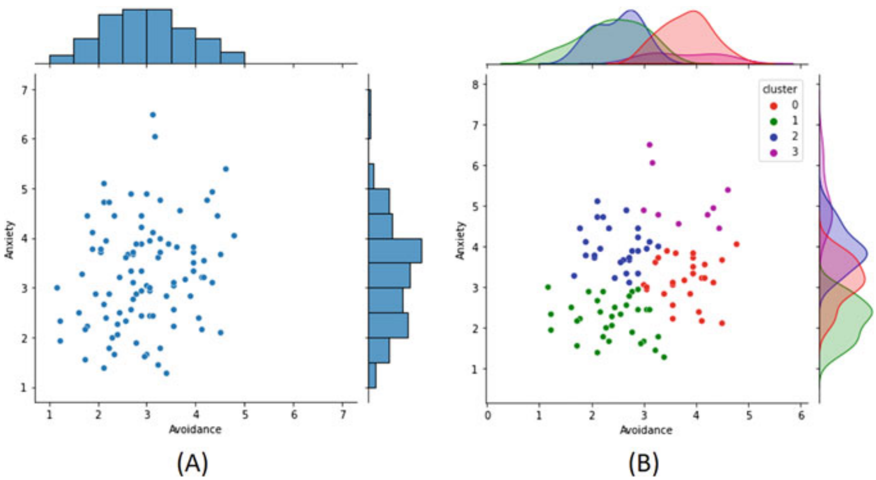


Fig. 1 Attachment results of the ECR-R questionnaire. **a** Unclustered results, **b** Clustered results

Noteworthily, the clustering results shown in the scatterplot are corroborated by the distribution of attachment styles in the population [19].

Stage 2—EEG Recording While Performing Success and Failure Tasks

Twenty-seven participants out of the 36 that were initially recruited participated in the second stage of the study. The distribution of this study population into attachment styles was as follows: 7 belonged to the avoidant group (#0), 6 belonged to the secure group (#1), 9 belonged to the anxious group (#2) and 5 belonged to the fearful avoidant group (#3).

In this stage, the players were engaged in performing the flanker task [20]. The flanker task is a widely used cognitive psychology experiment that measures the ability of an individual to focus attention on a specific target stimulus while ignoring distractors in their immediate environment. The flanker task is used to study the mechanisms of attention, cognitive control, and response inhibition. It has been employed in various fields such as neuroscience, clinical psychology, and cognitive aging research. One of four possible configurations of the arrow's flanker task (see Fig. 2) was presented for a duration of 1 s. The players were instructed to indicate the direction of a centrally located arrow (the target) flanked by non-target stimuli by pressing the corresponding right or left arrow on the keyboard. Each player was presented with 60 flanker trials divided into 3 blocks of 20 trials each. In the first and third blocks, the player had to press the keyboard arrow that was congruent with the direction of the target arrow. However, in the second block, participants had to press the keyboard arrow pointing to the opposite direction of the target. Feedback was provided at the end of each trial by changing the color of the feedback message that appeared on a subsequent slide. For correct trials the word "correct" appeared in green, whereas for incorrect trials the word "incorrect" appeared in red for a duration of 1 s. In between trials, participants were asked to focus their gaze at a gray cross situated in the middle of a black screen, for a random duration of 0.5 to 1.5 s, uniformly distributed.

The EEG signals were acquired by a 16 channels active EEG amplifier (g.USBAMP, by g.tec, Austria) with a sampling frequency of 512 [Hz] according to the 10–20 international system. Electrode impedance was kept under 5 [Kohm] during the entire experiment and was monitored by the OpenVibe [21] processing and recording software. Participants underwent a training session which allowed them to get familiar with the experimental procedure.

Fig. 2 Possible flanker task screens

3 Results

In order to remove noise and artifacts and enhance the EEG signal quality, we implemented a pre-processing pipeline as in previous studies (e.g. [22, 23]). The pipeline consisted of filtration (band pass [1, 32] Hz, combined with a notch filter at 50 Hz), re-referencing to the average reference and ICA decompositions. Finally, EEG epoching was performed with epoch duration = 1 s from the onset of each feedback slide. We focused our analysis only on the frontal and prefrontal electrodes (Fp1, F7, Fp2, F8, F3 and F7), due to the known frontal lobe involvement in cognitive processing (e.g., [24–27]).

In the next step, we evaluated each EEG epoch's complexity level by using Lempel–Ziv complexity (LZC) (e.g. [10, 28]) with eight bins. LZC measures the complexity of a sequence of symbols, such as an EEG epoch. LZC calculates the number of distinct patterns, in this case, quantized to 8 levels, that appear in a sequence, which can be used to estimate the amount of information contained in the sequence.

A two-way ANOVA was performed to analyze the effect of attachment style (secure or insecure) and feedback type (success or failure) on EEG signal complexity. The results revealed a statistically significant interaction between attachment style and feedback type ($F(1, 1596) = 6.483, p < 0.001$). Also, simple main effects analysis showed that attachment style and feedback type each had a statistically significant effect on EEG signal complexity ($p < 0.01$). The interaction effect is presented in Fig. 3.

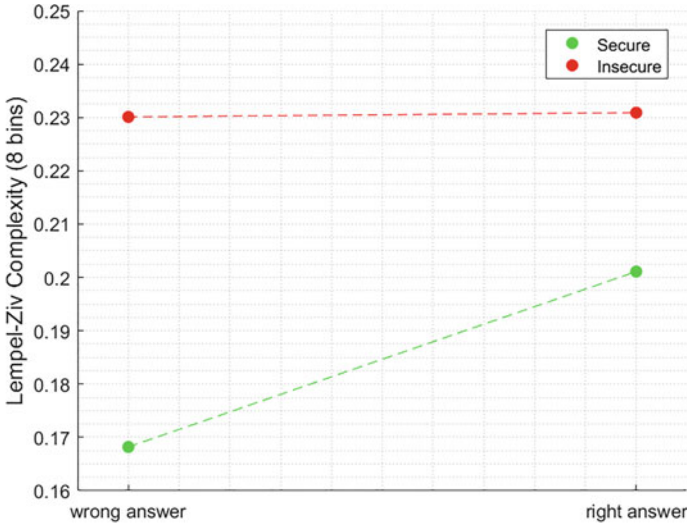


Fig. 3 The interaction between attachment style (secure or insecure) and feedback type (success or failure)

4 Conclusions and Future Work

The main goal of this study was to explore the effect of attachment style and feedback type on EEG signal complexity. For that purpose, we have designed a two-stage experiment. In the first stage, participants filled out an attachment style questionnaire, and those participants that were the best exemplars of secure or insecure attachment also participated in the second stage of the study where they performed a flanker task with feedback while EEG was simultaneously recorded from their scalp. The effect of attachment style (secure, insecure) and feedback type (success or failure) on signal complexity was measured by the LZC index.

The significant interaction between attachment style and feedback type was due to the change in complexity level between success and failure as a function of attachment type. Secure players were associated with an increase in complexity level between success and failure, whereas for insecure players no change was observed between these conditions. Additionally, insecure players were associated with a higher signal complexity level compared to secure players regardless of feedback type. Overall, these results suggest that insecure players may use a defensive mechanism against negative feedback, and may hint at different mechanisms of emotional regulation that are employed by secure and insecure participants.

There are several avenues for future research. First, additional features other than the complexity of the EEG should be investigated, such as frequency-based [29, 30], spatial-based [31], or time domain-based features [32]. Second, as the current study focused on frontal and prefrontal channels only, it would be worthwhile to use a denser network of electrodes which will allow to examine possible interactions with scalp distribution and perhaps allow the creation of spatial maps of activity. Third, it will be interesting to divide the insecure group into three subgroups, i.e., the anxiously attached group, the avoidant group, and the fearful avoidants/disoriented group, and check the statistical distribution of the various electrophysiological features depending on this distribution among all the subjects of the Insecure group. Finally, it would be interesting to create prediction models of all four attachment styles based on electrophysiological features by using machine learning methods.

References

1. Fearon, R. P., & Roisman, G. I. (2017). Attachment theory: Progress and future directions. *Current Opinion in Psychology, 15*, 131–136.
2. Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52*, 511.
3. Cassidy, J., & Shaver, P. R. (2002). *Handbook of attachment: Theory, research, and clinical applications*. Rough Guides.
4. Freeman, H., & Brown, B. B. (2001). Primary attachment to parents and peers during adolescence: Differences by attachment style. *Journal of Youth and Adolescence, 30*, 653–674.
5. Crowell, J. A., Treboux, D., & Waters, E. (1999). The adult attachment interview and the relationship questionnaire: Relations to reports of mothers and partners. *Personal Relationships, 6*, 1–18.
6. George, C., Kaplan, N., & Main, M. (1996). Adult attachment interview.
7. Feeney, J. A., Noller, P., & Hanrahan, M. (2014). Attachment style questionnaire. *Journal of Social and Personal Relationships.*
8. McIntosh, A. R., Kovacevic, N., & Itier, R. J. (2008). Increased brain signal variability accompanies lower behavioral variability in development. *PLoS Computational Biology, 4*, e1000106.
9. Ramadoss, J., Dawi, N. M., Rajagopal, K., & Namazi, H. (2022). Complexity and information-based analysis of the electroencephalogram (EEG) signals in standing, walking, and walking with a brain–computer interface. *Fractals, 31.*
10. Ibáñez-Molina, A. J., Iglesias-Parro, S., Soriano, M. F., & Aznarte, J. I. (2015). Multiscale Lempel-Ziv complexity for EEG measures. *Clinical Neurophysiology, 126*, 521–548.
11. Abásolo, D., Hornero, R., Gómez, C., García, M., & López, M. (2006). Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Medical Engineering & Physics, 28*, 315–322.
12. Hosseini, S. A., Khalilzadeh, M. A., Naghibi-Sistani, M. B., & Homam, S. M. (2015). Emotional stress recognition using a new fusion link between electroencephalogram and peripheral signals. *Iranian Journal of Neurology, 14*, 142.
13. Chen, T., Ju, S., Ren, F., Fan, M., & Gu, Y. (2020). EEG emotion recognition model based on the LIBSVM classifier. *Measurement, 164.*
14. Vrtička, P., & Vuilleumier, P. (2012). Neuroscience of human social interactions and adult attachment style. *Frontiers in Human Neuroscience, 6.*
15. Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*, 350.
16. Ridderinkhof, K. R., Wylie, S. A., van den Wildenberg, W. P. M., Bashore, T. R., & van der Molen, M. W. (2021). The arrow of time: Advancing insights into action control from the arrow version of the Eriksen flanker task. *Attention, Perception & Psychophysics, 83*, 700–721.
17. Jain, A. K. (2008). Data clustering: 50 years beyond K-means. In *Machine learning and knowledge discovery in databases*. Elsevier B.V.

18. Kodinariya, T. M. (2013). Review on determining number of cluster in K-means clustering. *International Journal of Advance Research in Computer Science and Management Studies, 1*, 90–95.
19. Magai, C., Cohen, C., Milburn, N., Thorpe, B., McPherson, R., & Peralta, D. (2001). Attachment styles in older European American and African American adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 56*, S28–S35.
20. Brunetti, M., Zappasodi, F., Croce, P., & Di Matteo, R. (2019). Parsing the Flanker task to reveal behavioral and oscillatory correlates of unattended conflict interference. *Science and Reports, 9*, 1–11.
21. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., & Lécuyer, A. (2010). Openvibe: An open-source software platform to design, test, and use brain–computer interfaces in real and virtual environments. *Presence Teleoperators & Virtual Environments, 19*, 35–53.
22. Mizrahi, D., Zuckerman, I., & Laufer, I. (2023). The effect of social value orientation on theta to alpha ratio in resource allocation games. *Information, 14*, 146.
23. Mizrahi, D., Zuckerman, I., & Laufer, I. (2022). Electrophysiological features to aid in the construction of predictive models of human-agent collaboration in smart environments. *Sensors, 22*, 6526.
24. Mizrahi, D., Laufer, I., Zuckerman, I. (2021). Topographic analysis of cognitive load in tacit coordination games based on electrophysiological measurements. In *NeuroIS Retreat 2021, Vienna, Austria.*
25. Gartner, M., Grimm, S., & Bajbouj, M. (2015). Frontal midline theta oscillations during mental arithmetic: Effects of stress. *Frontiers in Behavioral Neuroscience, 9*, 1–8.
26. Boudewyn, M., Roberts, B. M., Mizrak, E., Ranganath, C., & Carter, C. S. (2019). Prefrontal transcranial direct current stimulation (tDCS) enhances behavioral and EEG markers of proactive control. *Cognitive Neuroscience, 10*, 57–65.
27. Laufer, I., Mizrahi, D., & Zuckerman, I. (2022). An electrophysiological model for assessing cognitive load in tacit coordination games. *Sensors, 22*, 477.
28. Aboy, M., Hornero, R., Abasolo, D., & Alvarez, D. (2006). Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering, 53*, 2282–2288.
29. Mizrahi, D., Zuckerman, I., & Laufer, I. (2022). Analysis of alpha band decomposition in different level-k scenarios with semantic processing. In *Brain informatics: 15th international conference, BI 2022* (pp. 65–73). Springer International Publishing.
30. Mizrahi, D., Laufer, I., & Zuckerman, I. (2021). Level-K classification from EEG signals using transfer learning. *Sensors, 21*, 7908.
31. Zuckerman, I., Mizrahi, D., & Laufer, I. (2022). EEG pattern classification of picking and coordination using anonymous random walks. *Algorithms, 15*, 114.
32. Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neuroscience.*

The Effects of Webpage Prototypicality, Aesthetics, and Complexity on Eye Fixations and Company Perception

Aliaksei Miniukovich, Kathrin Figl, and Christiane Ernst

Abstract Webpage design affects both eye movements while viewing a page, and the quickly formed first impression, which in turn influences user preference and judgments of website and organizational goodness. We conducted an eye-tracking user study (n = 32) that revealed webpage properties—visual aesthetics, complexity, and prototypicality—to affect the number of fixations on a page, probably due to visually appealing content attracting the eye, more content or more confusing content requiring extra fixations, and more typical layout facilitating webpage exploration. Higher aesthetics consistently influenced user preference, resulting in more positive judgments regarding various aspects of organizational goodness, including perceived employer attractiveness, organizational competence, and purchase intention. Higher complexity negatively affected user preference and lowered perceived organizational goodness, but only for apparel-shopping websites and not for commercial banking. Higher prototypicality was associated with higher perceived organizational competence. An analysis of eye fixations on various webpage components suggested that user impressions may form based on the overall webpage design rather than any specific webpage element or component.

Keywords Prototypicality · Webpage design · User attitudes · Eye-tracking

A. Miniukovich (✉) · K. Figl · C. Ernst
Department of Information Systems, Production and Logistics Management, University of Innsbruck, Innsbruck, Austria
e-mail: miniukovich@e.atw-lab.com

K. Figl
e-mail: kathrin.figl@uibk.ac.at

C. Ernst
e-mail: christiane.ernst@student.uibk.ac.at

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_24

1 Introduction

The user perceives the appearance of a webpage very quickly [1], forming an initial impression of the webpage, website, and organization associated with the website [2]. This impression persists over time [3], partially resists a possible override by factual arguments (e.g., facts about the company's financial performance or stories about the treatment of employees), and influences users' preferences and actual behavior, such as product choice [4] and liking for the company [5]. Eye-movement patterns could help quantify and explain webpage perception and user preferences (cf., [6]), as eye movements are indicative of both the properties of webpages (e.g., visual complexity [7]) and mental processes (e.g., users like what they have looked at for longer [8]).

Past studies explored visual aesthetics [9], complexity [10], and prototypicality [11], and their relationships with gaze behavior separately, but not all three in the same study and in the theoretical context of the elaboration likelihood model (ELM). This paper presents a study that relied on ELM and concurrently explored webpage aesthetics, complexity, and prototypicality, while also analyzing user eye fixations to gain additional insights; Fig. 1 summarizes the main concepts and their relationships. Overall, the results confirmed that (a) visual features significantly impact user judgment of company goodness (attractiveness as an employer, organizational competence, and intention to purchase from a company), (b) visual features influence the amount of webpage viewing (the number of eye fixations), and (c) webpage viewing affects user judgment. Message strength (factual arguments) did affect user judgment, consistent with ELM. However, the type of information processing (in-depth or superficial, induced by cognitive load) did not moderate the effects of message strength and visual features on user judgment, contrary to what might have been expected.

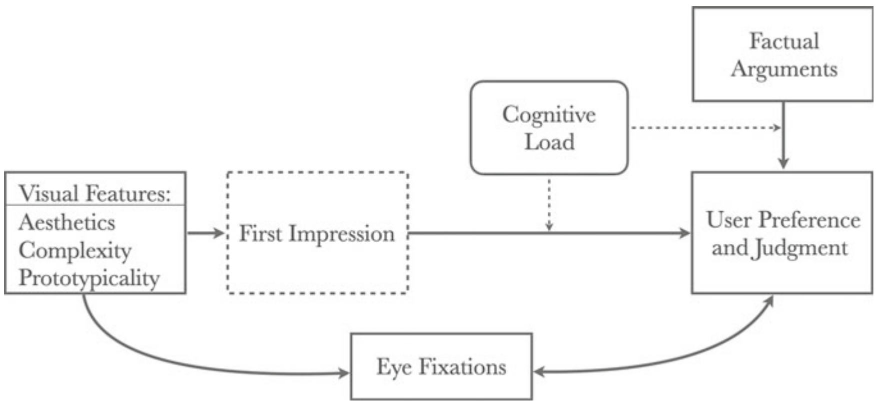


Fig. 1 A schematic representation of the main concepts studied

2 Related Work

Webpage Design Dimensions

Several design dimensions have been shown to influence user perception of webpages, including visual aesthetics [12] and complexity [13], design prototypicality [14], and perceived ease of use [15]. These dimensions are easily and almost immediately perceived by the mind [1], and their effects extend from a user's initial impression all the way to their overall evaluation of webpage quality, ultimately affecting their judgment of the organization behind the website [3]. The judgment itself relies on a variety of organization-related beliefs and intentions, such as the intention to purchase from or return to an online store [3], the belief in an organization's competence [16], or the attitude toward an organization as an employer [5].

The Elaboration Likelihood Model (ELM, [17, 18]) provides a theoretical explanation of the link between design dimensions and organization-related judgment: design dimensions act as cues of design quality—and, by association, of organization quality (cf., [12]). The users frequently rely on such task-unrelated cues for judgments, particularly when cognitive resources for processing webpages are limited, such as when users are under high cognitive load, bored, distracted, or have too many options to choose from. The task-related facts do affect judgment, but their effect diminishes as cognitive resources decrease and may still not be entirely independent of the task-independent but easily perceivable visual cues [17].

Unlike visual aesthetics (a webpage property that elicits a reasoning-free immediate pleasing feeling [19]) and complexity (a property of a webpage corresponding to the amount of mental effort required to comprehend a webpage), webpage prototypicality—how closely a webpage resembles the stereotypical webpage (for a given domain, e.g., an online shopping homepage or commercial banking homepage) in the user's mind—was only rarely included in studies of design effects on user preference or judgment. However, prototypicality, like aesthetics and complexity, appears to act as a peripheral cue and signal website and organizational quality to the user, resulting in, for example, higher purchase intentions [20]. Initial studies in HCI have also shown that higher prototypicality improves user performance in menu-exploration tasks [21, 22] and target-search tasks [23], as well as reduces confusion, which users are likely to appreciate and attribute to website quality.

Eye-Tracking

Eye-tracking is a popular technique for recording and analyzing how users view a webpage, providing insights into which parts of the page captured their attention and were explored in detail [6]. The human eye receives most of its incoming information from only a small central part of the visual field, known as the fovea. The mind directs the fovea towards areas of interest that are either task-relevant or visually conspicuous. Fixations are periods of no eye movement during which the mind extracts visual information from the area that the fovea covers (called a fixation location), lasting approximately 250ms on average [24]. Rapid eye movements from one fixation location to another are called saccades, during which the mind perceives no information.

Past research found that gaze behavior depends on visual stimuli properties such as aesthetics, complexity, and prototypicality. For instance, aesthetics correlated with dwell time for images of cars [7] and lamps [9], possibly because participants preferred to look at nicer images for longer in tasks with no time constraints. However, other research on simple-shape stimuli [8] suggested that looking per se, rather than aesthetics, was a source of preference on its own, with the user seemingly choosing the stimuli they looked at the longest. Paying attention to a specific option in a set strongly correlates with the likelihood of being chosen, as seen in consumer decisions [25].

Visual complexity has also been found to correlate with the number of fixations in free-viewing tasks [7], possibly due to complex stimuli having more interest points to look at and process [26]. However, the relationship between complexity and the number of fixations may not be linear and may be moderated by task complexity and cognitive load. Simple tasks may allow for more exploration of complex webpages and, therefore, more fixations, whereas complex tasks may limit user fixations to task-relevant areas, resulting in fewer fixations [10], which is consistent with the load theory of attention [27].

Similar to aesthetics and complexity, design prototypicality has been found to correlate with the number of fixations, as observed in studies of pastry products [11], and this correlation then translated into purchase behavior, with the prototypicality-purchasing relationship disappearing after accounting for fixations. However, the effects of cognitive load on the prototypicality-purchasing relationship, as well as its interaction with fixations, have not yet been explored.

Gaze behavior on webpages is distinct from gaze behavior on simpler stimuli, such as shape patterns, which are commonly used in cognitive research. The user mind, for example, divides webpages into semantically coherent chunks (groups of related elements) and allocates attention and eye fixations to these chunks rather than individual items [28]. An in-depth examination of fixations on various semantic chunks (e.g., webpage header and footer, menus, or large-image rotating banners) may reveal which parts of webpages have the greatest impact on user preference and behavior.

3 Method

We conducted an in-lab eye-tracking study to investigate the effects of design variables—specifically, webpage prototypicality, aesthetics, and complexity—on company perception and judgment. Even though the task scenario focused on job searching, the study had participants rate organizational attractiveness as an employer, organizational goodness, and purchase intention to cover a range of judgments related to company goodness. The study also investigated the effects of design variables on eye movements while viewing a webpage, as well as the effects of fixating on various webpage elements on user judgment.

Design

The study used a mixed design, with cognitive load (low and high) as a between-subject factor and webpage prototypicality (low and high) and message strength (low and high) as within-subjects factors.

Stimuli

The study utilized a total of 48 webpages—24 from commercial banks (12 low and 12 high prototypicality) and 24 from apparel shopping sites (12 low and 12 high prototypicality)—which were randomly paired with 48 organization-related messages for each participant. The messages included 24 low and 24 high message strength messages, with half being "About Us" company descriptions and half being company reviews by employees, each with a mean length of 71.3 words. The homepages were selected from a larger pool of 1032 bank homepages and 550 apparel store homepages that had previously been pre-rated on visual complexity, aesthetics, and prototypicality in separate crowdworker-based studies: webpages from both extremes of the prototypicality rating distributions were screened for potential confounders (e.g., a bank's country of origin and the number of branches, or type of apparel product) and selected in the high- and low-prototypicality groups. The messages were also chosen from a larger pool of 68 messages that were pre-rated on message strength (using three items: "*The arguments of this message are convincing/strong/informative*").

Participants

We recruited 32 participants (17 female) by utilizing a dedicated emailing list from a local university ($M_{age} = 25.0$ years, $SD_{age} = 2.9$). Half of the participants were from Germany, while the other half consisted of participants from Austria (5), Italy (4), and other countries (7). Most of the participants were students, with only one participant being fully employed and another participant in a traineeship. Out of the student participants, most studied social sciences (22), with others studying engineering (5), natural sciences (2), or something else (2). Everyone in the study had normal color vision.

Procedure

After being provided with a study description and undergoing eye-tracker calibration (Tobii TX300), the participants completed a demographic questionnaire, read a scenario, and completed 16 trials. The 16 webpages (eight banks and eight apparel shopping sites, four each with high prototypicality and four with low prototypicality) were randomly selected per participant, with care taken to ensure that all 48 webpages received an equal number of ratings. During each trial, participants first viewed a pattern that they had to memorize for five seconds (the high cognitive load group had to remember five cells within a four-by-four grid, while the low load group had to remember two cells), read a message, viewed a screenshot of a homepage (1440px wide, full-page length, centered, scrollable) without a time limit, and then rated an organization. As the final step, they reproduced the memorized pattern. A session lasted 35–40 min, and participants were paid ten euros for their time. Participants rated their intention to apply for a job at an organization on items using a seven-point scale (three items, adapted from [29], sample item: "*If I were offered a job at this organization, I would accept it*", "*Definitely not*" to "*Definitely would*"), organization competence (three items, adapted from the 'ability' scale [30], sample Likert-type item "*The organization provides good services*"), and intention to purchase (three items, cf., [3], sample item "*I would become a customer/client of this organization if I needed the type of services they offer*", "*Very likely*" to "*Very unlikely*").

Scenario

Participants were given the following task scenario: "*Imagine that you are looking for an assistant-manager job and have narrowed down the list of potential employers to a few organizations that suit you location-wise and have openings. Consider each organization individually, and rate your preference for employment with each.*"

Classification of Fixations

We classified each fixation as landing on a web primitive (text, image, or clickable UI element, Fig. 2) or semantic element (a header, footer, main content, menus, interactive forms, section titles, rotating banners, or social media areas) to investigate the effects of different elements on webpage perception. First, the fixations recognized by the eye-tracker were cleaned of any fixations that landed outside of

Fig. 2 Visualizations of gaze paths for three participants, overlaid on different web primitives: images (red), texts (yellow), and functional UI elements (green). The last participant (far right) appears to have explored the webpage superficially, with only a few fixations

the webpage screenshot. Second, fixation location coordinates were mapped onto webpage screenshot coordinates, with fixations that occurred during scrolling being filtered out (although the parts of the fixation that occurred before scrolling started were retained as separate fixations). Finally, each remaining fixation was automatically classified based on the element or primitive it landed on. If it did not land on any specific element, it was categorized as "other." The coordinates of elements and primitives were also extracted automatically.

4 Results

A review of fixation histograms revealed that several semantic elements had few fixations (forms, titles, banners, and social media). Therefore, their fixations were converted from counts to binary variables. With the exception of images, the fixation count distributions were positively skewed and were subsequently log-normalized. Since dwell time was redundant to fixation counts, it was dropped from the analysis.

A series of linear mixed models were run for each of the three dependent variables (job attractiveness, organizational competence, and purchase intention), with participant ID, webpage ID, and message ID as random effects (intercepts only), domain (fashion vs. banks), and cognitive load as moderators, and message strength and page height as extra fixed effects to control for. Fixation counts for each of the primitives or semantic elements were predictors (entered one at a time). Our results indicated that only fixation counts for textual and UI primitives, as well as for the webpage's main body and menus, along with the total number of fixations on the entire page (Table 1), had a significant effect on the dependent variables. Cognitive load did not affect the dependent variables and did not interact with fixations. A closer examination of the fixations on controls (UI) and menus in the same model revealed that they were redundant. When fixation counts were combined with entire-page fixation counts in the models, their effects became insignificant, and we retained only the entire-page fixation counts for further analyses. Finally, our findings revealed that message strength always had a significant effect on the dependent variables.

A series of mixed models with the entire-page fixation counts as the dependent variable (Table 2) revealed that all three design variables increased the counts, as expected, as did the page height. On top of fixation counts, aesthetics consistently had a positive effect on the three dependent variables, whereas higher prototypicality only positively influenced perceptions of organizational competence, and complexity negatively affected all dependent variables, but only for apparel (Table 3).

5 Discussion

The results generally aligned with expectations, with the three design variables having significant effects on the dependent variables: higher aesthetics had a positive impact on all three dependent variables, while complexity had a negative effect on all three, but only for the apparel domain. In contrast, prototypicality had a positive effect only on perceived organizational competence (Table 3). Interestingly, for the apparel domain, higher prototypicality led to a decrease in purchase intention, possibly indicating that users value novelty in fashion shopping but safety in online banking. In line with ELM, message strength affected all three variables, but its effect was consistent under both high and low cognitive load, which was unexpected. We can speculate that either a more challenging, concurrent task is required (e.g., listening to a sequence of words and counting nouns while viewing a webpage), or that users perceive web design quality as a central, non-peripheral cue.

Fixation counts were also influenced by all three design variables (Table 2). Higher aesthetics led to a greater number of fixations, supporting the notion that users spend more time looking at aesthetically pleasing designs. Moreover, the effect of aesthetics on user preferences (Table 3) remained significant even after accounting for fixations, which were significant predictors of employer attractiveness and purchase intention, implying the existence of biases for both fixated-on objects and aesthetically pleasing objects (cf., [8]). Complexity increased fixation counts, consistent with webpage complexity being a proxy for "more items to look at" (cf., [26]), but did not interact with load, as expected from the load theory [27]. Higher prototypicality also increased fixation counts, suggesting that either prototypicality acted through

Table 1 Linear mixed models (each row is a model) of employer attractiveness (JOB), organizational competence (COMP), and purchase intention (PURC), with fixation counts (Fix), cognitive load (Load, high load as a baseline), domain (Dmn, Banks as a baseline), message strength (Msg, high strength as baseline), and page height (H) as fixed effects. R^2 indicates an improvement in fit over the random effects-only model. Significant predictors ($p < 0.05$) are highlighted in gray *

Output	Fixations on	Beta coefficients								R^2
		Fix.	Fix:Load	Load	Fix:dmn	Dmn	Msg	Msg:Load	H	
JOB	Texts	.17	-.11	-.13	-.05	-.06	-.32	-.14	.10	.07
	Controls	.18	-.04	-.15	-.02	-.16	-.34	-.11	.10	.07
	Main body	.15	-.10	-.17	.14	-.05	-.34	-.13	.06	.08
	Menus	.11	.09	-.14	-.04	-.22	-.33	-.13	.15	.08
	Entire page	.17	-.08	-.19	.08	-.07	-.34	-.11	.07	.08
COMP	Texts	.13	.12	.00	-.16	.11	-.32	-.03	.20	.07
	Controls	.16	-.01	.01	-.12	-.00	-.34	-.02	.24	.06
	Main body	.14	.03	-.01	-.08	.08	-.34	-.02	.19	.06
	Menus	.15	-.04	.02	-.12	-.02	-.33	-.02	.26	.06
	Entire page	.12	.07	-.02	-.09	.06	-.34	-.01	.21	.06
PURC	Texts	.20	.04	.31	-.18	.37	-.17	-.16	.15	.07
	Controls	.18	.06	.30	-.12	.22	-.18	-.24	.18	.07
	Main body	.18	.04	.27	-.05	.34	-.18	-.15	.12	.07
	Menus	.18	.06	.32	-.17	.17	-.18	-.16	.22	.07
	Entire page	.18	.11	.26	-.10	.31	-.18	-.13	.13	.08

Table 2 A series of linear mixed models predicting entire-page fixation counts

Design Variable	Beta coefficients						R^2
	DesV	DesV:Load	Load	DesV:Dmn	Dmn	H	
Aesthetics	.21	.00	.34	-.19	-.40	.37	.10
Complexity	.28	.04	.35	-.24	-.27	.30	.11
Prototyp.Gr.	.24	-.14	.28	.03	-.43	.44	.09

aesthetics [11] or that users prefer to look at items that feel safer and more familiar, which prototypicality may correspond to [31].

Fixations on several webpage components influenced user preference (Table 1). Surprisingly, fixation on images did not have a significant effect, contrary to expectations, as higher numbers of images typically correspond to higher web page aesthetics [32]. This finding is consistent with past studies indicating that few fixations land on images during goal-oriented viewing [33], although the reasoning behind participants not fixating on task-unrelated images may not apply to our free viewing task. None of the counts of fixations on specific elements or primitives had a stronger influence on user preference than the counts for the entire page. This implies that either the entirety of a webpage affects the user rather than its specific components, or that the user explores a webpage along a pre-defined, possibly learned pathway that includes looking at most of non-image parts of the webpage.

6 Limitations

The models (Table 3) only explained a modest amount of variance in user judgments of organizational goodness, and future studies may need to include other predictors, either design-based (such as the amount and quality of textual information) or demographic (such as individual disposition towards different business domains). Future research may also need to incorporate goal-oriented tasks, such as searching for a specific piece of information, because some of our participants may have taken advantage of the free-viewing task and viewed webpages too briefly (resulting in very few fixations) to be realistically influenced by their design.

Finally, larger studies with more participants may be necessary to re-test whether some of our insignificant results were caused by Type 2 errors. For example, with 134 participants, the insignificant effect of on-image fixation counts on purchase intention (beta coef. = 0.13, when in a model similar to those in Table 1) could become significant (power = 81%, CI.95 = 71.93–88.16). Similarly, the interaction between message strength and cognitive load (eff. size = −0.16, Table 3, first row) that was expected based on the ELM but remained unobserved in our study might have been significant (power = 81%, CI.95 = 71.93–88.16) in a study with 156 participants. Therefore, future research with a larger sample size may help confirm or refute some of the non-significant findings reported in this study.

Table 3 Linear mixed models, with the three design variables (DesV; high/low proto for DesV:Load interaction), entire-page fixation counts (Fix), and other variables as fixed effects

Output	Design Variable	Beta coefficients								R^2
		DesV	DesV:Load	Fix.	DesV:Dmn	Dmn	Msg	Msg:Load	H	
JOB	Aesthetics	.21	-.07	.14	-.10	-.10	-.31	-.16	.03	.09
	Complexity	.14	-.00	.15	-.33	-.15	-.31	-.17	.14	.09
	Proto.Gr.	.23	-.11/-.25	.17	-.25	.06	-.33	-.12	.06	.08
COMP	Aesthetics	.44	-.16	.08	-.29	.06	-.34	-.01	.11	.10
	Complexity	.25	.12	.09	-.50	.05	-.33	-.02	.21	.09
	Proto.Gr.	.48	.01/-.05	.10	-.36	.24	-.34	-.01	.20	.07
PURC	Aesthetics	.33	-.02	.17	-.21	.30	-.24	-.01	.06	.09
	Complexity	.28	.16	.17	-.56	.32	-.24	-.01	.13	.11
	Proto.Gr.	.40	.26/.25	.18	-.53	.60	-.19	-.12	.11	.08

7 Conclusion

The paper examined the effects of webpage design on user perception of an organization, using eye movements as a relevant descriptor of both design features and perception. The study was novel in investigating the effect of one design variable—prototypicality—while also corroborating past findings on the impact of webpage aesthetics and complexity on website and organization perception. Our observation of both aesthetics and fixation counts being significant (Table 3) contributes to the ongoing debate about the nature of preference—whether people prefer to look at nicer things or like things they look at more [8]—implying that both principles apply. Finally, our investigation into the relative contributions of different webpage elements to webpage perception (Table 1) was also novel, as previous research had primarily focused on specific features (e.g., images of human faces [34]) rather than on the contributions of different element types and groups. Our results suggest that overall fixation counts may explain appreciation and judgment better than individual counts for different element types. Overall, this paper sheds light on the complex relationship between webpage design and user perception, highlighting the utility of multiple visual variables to webpage design practice, and eye movement patterns to user experience practice.

Acknowledgements The M-2827 project grant of the FWF (Austrian Science Fund) supported this research.

References

1. Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology, 25*, 115–126.
2. Kim, H., & Fesenmaier, D. R. (2008). Persuasive design of destination websites: An analysis of first impression. *Journal of Travel Research, 47*, 3–13.
3. Thielsch, M. T., Blotenberg, I., & Jaron, R. (2014). User Evaluation of websites: From first impression to recommendation. *Interacting with Computers, 26*, 89–102.
4. Deng, W., & Chen, L. (2022). Do first impressions matter? Effect of first impressions on tourists' destination choice: an event-related potential study. *Current Issues in Tourism, 26*, 1518–1535.
5. Braddy, P. W., Meade, A. W., & Kroustalis, C. M. (2008). Online recruiting: The effects of organizational familiarity, website usability, and website attractiveness on viewers' impressions of organizations. *Computers in Human Behavior, 24*, 2992–3001.
6. Djamasbi, S. (2014). Eye Tracking and Web Experience. *AIS Transactions on Human-Computer Interaction, 6*, 37–54.
7. Chassy, P., Lindell, T. A., Jones, J. A., & Paramei, G. V. (2015). A relationship between visual complexity and aesthetic appraisal of car front images: An eye-tracker study. *Perception, 44*, 1085–1097.
8. Isham, E. A., & Geng, J. J. (2013). Looking time predicts choice but not aesthetic value. *PLoS ONE, 8*, 1–7.

9. Guo, F., Li, M., Hu, M., Li, F., & Lin, B. (2019). Distinguishing and quantifying the visual aesthetics of a product: An integrated approach of eye-tracking and EEG. *International Journal of Industrial Ergonomics, 71*, 47–56.
10. Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems, 62*, 1–10.
11. Scarpi, D., Pizzi, G., & Raggiotto, F. (2019). The extraordinary attraction of being ordinary: A moderated mediation model of purchase for prototypical products. *Journal of Retailing and Consumer Services, 49*, 267–278.
12. Tractinsky, N., & Lowengart, O. (2007). Web-store aesthetics in e-retailing: A conceptual framework and some theoretical implications. *Academy of Marketing Science Review, 11*, 1–18.
13. Deng, L., & Poole, M. S. (2010). Affect in web interfaces: a study of the impacts of web page visual complexity and order. *MIS Quarterly, 34*, 711–730.
14. Tuch, A. N., Presslaber, E. E., Stöcklin, M., Opwis, K., & Bargas-Avila, J. A. (2012). The role of visual complexity and prototypicality regarding first impression of websites: working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies, 70*, 794–811.
15. Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: effects on user performance and perceived usability. *Applied Ergonomics, 41*, 403–410.
16. Connolly-Ahern, C., & Broadway, S. C. (2007). The importance of appearing competent: An analysis of corporate impression management strategies on the world wide web. *Public Relations Review, 33*, 343–345.
17. Samson, A., & Voyer, B. G. (2012). Two minds, three ways: dual system and dual process models in consumer psychology. *AMS Review, 2*, 48–71.
18. Kitchen, P. J., Kerr, G., E. Schultz, D., McColl, R., & Pals, H. (2014). The elaboration likelihood model: review, critique and research agenda. *European Journal of Marketing, 48*, 2033–2050.
19. Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies, 68*, 689–709.
20. Celhay, F., & Trinquecoste, J. F. (2015). Package graphic design: Investigating the variables that moderate consumer response to atypical designs. *Journal of Product Innovation Management, 32*, 1014–1032.
21. McCarthy, J. D., Sasse, M. A., & Riegelsberger, J. (2004). Could I have the menu please? An eye tracking study of design conventions. In: *People and computers xvii—Designing for society* (pp. 401–414). Springer.
22. Santa-Maria, L., & Dyson, M. C. (2008). The effect of violating visual conventions of a website on user performance and disorientation: How bad can it be? In: *International Conference on Design of Communication* (pp. 47–54). ACM.
23. Roth, S. P., Tuch, A. N., Mekler, E. D., Bargas-Avila, J. A., & Opwis, K. (2013). Location matters, especially for non-salient features–an eye-tracking study on the effects of web object placement on different types of websites. *International Journal of Human-Computer Studies, 71*, 228–235.
24. Galley, N., Betz, D., & Biniossek, C. (2015). Fixation durations-why are they so highly variable? In: *Advances in visual perception research* (pp. 83–106).
25. Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144*, 190–206.
26. Miniukovich, A., Sulpizio, S., & De Angeli, A. (2018). Visual complexity of graphical user interfaces. In: *International Conference on Advanced Visual Interfaces* (pp. 1–9).
27. Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*, 339–354.
28. Yesilada, Y., Jay, C., Stevens, R., & Harper, S. (2008). Validating the use and role of visual elements of web pages in navigation with an eye-tracking study. In: *International Conference on World Wide Web* (pp. 11–20).
29. Turban, D. B. (2001). Organizational attractiveness as an employer on college campuses: An examination of the applicant population. *Journal of Vocational Behavior, 58*, 293–312.

30. Brengman, M., & Karimov, F. P. (2012). The effect of web communities on consumers' initial trust in B2C E-commerce websites. *Management Research Review, 35*, 791–817.
31. Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science, 17*, 799–806.
32. Miniukovich, A., & Angeli, A. D. (2016). Webpage aesthetics: One size doesn't fit all. In: *Nordic Conference on Human-Computer Interaction.* Association for Computing Machinery.
33. Espigares-Jurado, F., Munoz-Leiva, F., Correia, M. B., Sousa, C. M., Ramos, C. M., & Faísca, L. (2020). Visual attention to the main image of a hotel website based on its position, type of navigation and belonging to millennial generation: An eye tracking study. *Journal of Retailing and Consumer Services, 52*, 101906.
34. Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring human images in website design: A multi-method approach. *MIS Quarterly, 33*, 539–566.

HUBII—Towards an Open Human Biosignal Intelligence Platform

Ivo Benke, Elias Mueller, and Alexander Maedche

Abstract With the rise of biosignal sensor technology and the wide adoption of wearables like smartwatches, the ability to collect high-quality human biosignals in the field has grown rapidly. Researchers may use these signals to investigate human states like emotions or workload and design advanced adaptive systems. However, processing biosignals is highly complex, limiting access to the significant potential of biosignals to domain experts only. To provide accessibility to biosignals for the NeuroIS community and researchers in general on a large scale, we present the concept of HUBII, the open Human Biosignal Intelligence platform. HUBII deploys biosignal processing, modelling, and consumption pipelines. It allows researchers to easily use these pipelines and contributes with making biosignal processing accessible, transparent through its open character, and existing work of other researchers reusable.

Keywords Biosignals · Processing · Modelling · Platform · NeuroIS

1 Introduction

With the rise of biosignal sensor technology and wearables like smartwatches or chest belts the ability to collect high-quality biosignals of humans has increased, not only in the lab but also in the field. Biosignals are autonomous signals produced by organisms, energetically measurable in physical quantities using sensors [1]. Body-attached wearables are able to capture electrical biosignals such as heart activity

I. Benke (✉) · E. Mueller · A. Maedche
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: ivo.benke@kit.edu

E. Mueller
e-mail: elias.mueller@kit.edu

A. Maedche
e-mail: alexander.maedche@kit.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_25

measured by electrocardiography (ECG), muscle activity recorded by electromyography (EMG), eye activity measured by electrooculogram (EOG), or skin conductance measured by electrodermal activity (EDA). Further, there are kinematic, optical, chemical, acoustic, and thermal biosignals which can also be measured via today's smartwatches through sensors. Collected biosignals, in combination with machine learning (ML) techniques, bear huge potential to recognize human traits and states, for example, personality, flow, cognitive workload, or emotions [2–5], and understand human behavior. Thereby, systems may become adaptive and support the user in productivity and well-being. This offers a unique opportunity for the field of NeuroIS for the design and application of innovative neuro-adaptive systems [5] (for example, passive brain-computer interfaces [6]).

While the ability to collect biosignals has grown rapidly, processing and analyzing them is highly complex and still poses multiple challenges. Depending on the specific sensor and its technical characteristics, data processing requires multiple steps. On the one hand, applying these steps in individual projects requires not only expert knowledge but is time-consuming, not trivial, and highly individual to each project. Consequently, despite the vast potential of biosignals, access to them and their usage on a large-scale is still restricted. As a result, this hinders the large-scale adoption of biosignal-based empirical research and design science research, e.g., designing user- and neuro-adaptive systems.

This challenge is, for example, similar to the accessibility gap of ML knowledge or natural language processing (NLP) in the past. While ML techniques provided significant potential and still do the access to it was often limited to experts from statistics, data, or computer science. To overcome this gap, automatic ML platforms (i.e., AutoML) arose, which provide access to ML modelling, training, and prediction without detailed code knowledge [7]. Similarly, the famous platform Hugging Face (www.huggingface.co) opened the closed domain of NLP techniques for users by provisioning the python-based transformers library. This gave easy access to NLP models and datasets on a large scale with a standardized interface. For biosignals, however, such a solution is not yet in sight.

We suggest providing this accessibility for the NeuroIS community and supporting standardized processing and analysis of human biosignals on a large scale for different applications. In this paper, we present a first architectural concept and a prototypical implementation of the Human Biosignal Intelligence (HUBII) platform. The platform consists of three tightly integrated building blocks, the biosignal processing, modelling, and consumption blocks. The HUBII platform may support IS researchers and related fields which are not deep experts by making biosignal processing accessible. Second, it makes processing and analysis transparent through its open and easily expandable character. Finally, it makes existing work reusable for others and promotes efficient processing solutions. With this paper, we want to stimulate a discussion in the NeuroIS community on the potential of such an "intelligence platform" supporting human biosignal processing on a large-scale.

2 The Open Human Biosignals Intelligence Platform HUBII

The overarching goal of HUBII is to become a platform for researchers to process and analyze biosignals as well as model and predict human traits and states based on these biosignals. The idea of HUBII is that researchers alike can process and analyze their collected biosignal data with very low effort, and either receive preprocessed "clean" features from their data or get predictions for specific human traits and states directly. Thereby, everybody has easy-to-use access to biosignal data processing and classification with HUBII which is transparent to the public at scale.

Figure 1 presents the architectural concept of the HUBII platform. At its core HUBII consists of three building blocks. The building blocks contain the processing and analysis of biosignal data as well as prediction of human traits and states. For the different biosignals, HUBII deploys one or multiple pipelines that are explicitly designed to run the specific data stream. Each of these pipelines is connected to one of the three fundamental building blocks: The *(1) Biosignal Processing* building block in which the raw data is processed into analyzable features, the *(2) Biosignal Modelling* building block in which modelling pipelines are hosted and models are trained, and the *(3) Biosignal Consumption* building block for providing services to prediction of human traits (e.g., personality) and states (e.g., emotions, workload).

In the following, we shortly outline the capabilities of HUBII along the three building blocks introduced above:

1. **Biosignal Processing**. This building block focuses on the processing of the biosignal raw data delivered from the sensor to derive relevant features from it. As depicted in Fig. 1 the biosignal raw data is provided by a specific sensor technology, e.g., an ECG in a smartwatch or chest strap. The data is sent to HUBII via a standardized API call. In the API call, the pipeline, the specific sensor, the

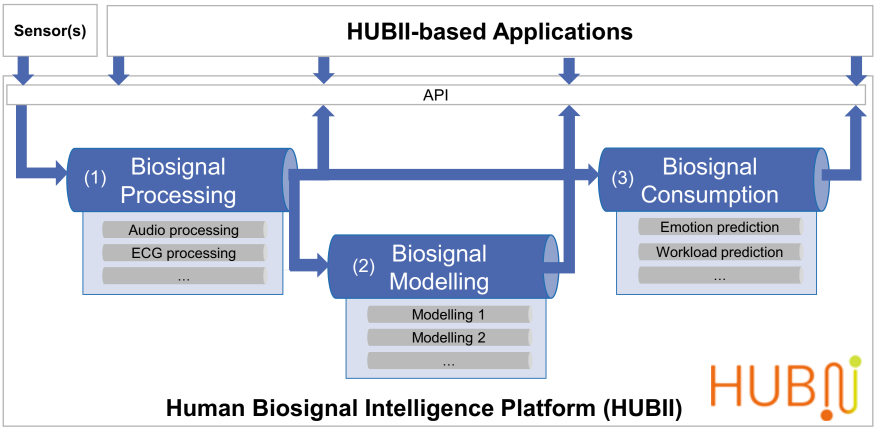


Fig. 1 Architectural concept of the HUBII platform with three building blocks

data format, and sensor specific parameters are defined. For example, for an ECG signal it would be required to articulate that the biosignal raw data is provided by a smartwatch (e.g., Samsung Galaxy Pro 5) in a corresponding data form (e.g., JSON). These predefined formats assure a sufficient data processing and quality level. Furthermore, the sliding window to derive the time-related features should be defined. Subsequently, the corresponding ECG pipeline processes the data, extracts the relevant features, and stores them in an appropriate format such as in a time series format for heartrate data. In addition, HUBII provides the ability to upload self-reported subjective labels for human traits and states (e.g., collected via experience sampling method) together with the biosignal raw data. This enables the creation of training data sets with labels as input for biosignal modelling.

2. **Biosignal Modelling**. The modelling building block hosts pipelines to create biosignal-based classification models for human traits and states based on collected training data. The training data not necessarily has to be processed by the biosignal processing building block of HUBII. It is also possible to upload existing training datasets that contain pre-processed features and relevant labels (with regards to the specific trait or state). Furthermore, it is worth mentioning that the proposed modelling pipelines support combining multiple biosignals following a multimodal approach. Subsequently, the modelling layer applies multiple supervised machine learning algorithms as part of the modelling pipeline. As a result, HUBII users receive the trained models and the evaluation metrics for them. Another feature of HUBII is that models can be uploaded to HUBII and subsequently hosted in the consumption building block as publicly available service.
3. **Biosignal Consumption**. The third building block of HUBII focuses on supporting consumption. Models for multiple human traits and states are hosted in pipelines. The consumption building block takes processed biosignal data in form of specific features as input and returns a value for a specific trait (e.g., personality) or state (e.g., emotion, workload). The resulting information can be used for multiple purposes in HUBII-based applications, e.g., to analyze human traits and states or design specific user-adaptive systems. The input features, thereby, do not have to belong to one biosignal only. For example, emotion can be predicted following a multimodal approach based on a fusion of ECG, acceleration, and EDA biosignals.

With HUBII, we intend to provide an integrated platform to process biosignals data towards analyzable features, allowing to model and develop robust classifiers, and providing seamless ways to receive predictions about human traits and states such as emotions or cognitive workload based on hosted ML models. With these functionalities, HUBII will provide key benefits to the NeuroIS community.

Accessibility. HUBII allows users to access biosignals processing, modelling, and consumption capabilities without deep knowledge on the programming steps in the respective building block. This allows a much larger group of users to leverage biosignals for understanding human behavior and giving it access to design advanced user- and neuro-adaptive systems. This may strongly accelerate the development of NeuroIS research.

Reuse. Today, many researchers interested in biosignals are developing individual code to process, model, or consume biosignal data depending on the biosignal sensor technology, the signal in use, and the human trait and state of interest. After the usage, no other researcher is likely using the developed code again which makes the development inefficient and fragmented. This in turn, prevents learning from previous mistakes and, thereby, hinders progress in research findings. On HUBII, a multitude of pipelines is hosted and reused. As beneficial effect, efficient and effective pipelines are used and represent a natural selection of high quality biosignal pipelines.

Transparency. The pipelines hosted on HUBII shall be transparent. That means that the steps each pipeline does will be documented in a transparent way to let researchers understand the corresponding process activities. Together, with the respective outputs of the pipelines and a clear description, there will be functions to receive a detailed list of conducted steps and the computational methods used. Following an open science approach [8, 9], as a result, researchers can refer to the pipeline as to proof good practices which are easily replicable and justifiable.

Empirical Research and Design Science Research Perspective on HUBII

For a better understanding on how to use the HUBII platform, we want to outline two different perspectives with a focus on empirical and design science research. These perspectives emphasize different capabilities of HUBII for different purposes:

1. **Empirical Research Perspective**. First, HUBII shall be a platform for empirical researchers who aim to investigate socio-technical phenomena leveraging biosignals. Due to the advantages of its accessible and transparent building blocks it is possible to leverage HUBI in different forms: First, in the biosignal processing building block they can derive analyzable features from their raw data. Afterwards researchers have two options. They can use the received features for an analysis in their empirical study. Alternatively, they can use the results of processed features to receive a prediction for specific human traits or states in the consumption building block. Depending on the target prediction they can use the results of multiple processing pipelines. For example, prediction of cognitive workload might be based on the ECG and EEG features. This is defined by the prediction pipeline used. Overall, the focus in empirical research is to leverage HUBII capabilities for processing biosignal data as part of empirical studies.
2. **Design Science Perspective**. Besides the capabilities to support analyzing biosignals and provide predictions for specific human traits and states, HUBII provides functionalities for researchers to finetune models to predict traits or states or create new models based on biosignal data. Furthermore, HUBII targets

supporting the design of user-adaptive systems that process biosignals in real-time and trigger system-generated adaptations automatically. This is the design science perspective of HUBII. Design science researchers may upload their collected signals (e.g., heart rate) in combination with labels to the respective pipeline. This is a key difference to the HUBII empirical research perspective. In the biosignal processing building block the corresponding pipeline derives features explicitly tailored depending on the uploaded labels. For example, heart rate features are based on the sliding windows defined by the labels and their definition. Subsequently, the modelling building block is intensively used by design science researchers. As results, researchers receive trained models and the evaluation metrics for the models. In a subsequent step, the models can be deployed in the consumption building block and support designing user-adaptive system on top of it.

Exemplary HUBII Platform Services

HUBII[1] is made available to the public to allow researchers to leverage its benefits. To highlight the potential of HUBII, we describe two exemplary services that we already provide on HUBII with a specific focus on the capabilities of the biosignal processing building block:

(1) **Heart Rate Variability (HRV) Service**. A usual task for research leveraging heart rate measurements is the derivation of features from the raw ECG signal. In this case, we provide a service through the biosignal processing building block for extracting time domain and frequency domain HRV features. In the current pipeline implementation, the raw signals of a Polar H10 chest belt sensor are taken as input. Since the data transmission rate of the Polar H10 sensor is 1 Hz, raw data does not need to be down-sampled. Raw data may be uploaded either via an integrated user interface in CSV format or via a programmable API request in JSON format. As a result, users receive the time domain and frequency domain HRV features of the input data in a JSON or CSV format. The HRV processing pipeline is available under www.hubii.world/processing/hrv-01.

(2) **Audio Diarization Service**. Audio is a rich biosignal by which humans transfer language and content. An important processing step required for audio signals with multiple speakers involved is the derivation of speaker assignment and corresponding text extraction. We provide this service out of the box via the HUBII biosignal processing building block. Users can upload an audio file (mp3, wav) with dialogues of individuals. The service segments the speakers' contributions (utterances) and transcribes the segments of the speakers into textual content. The resulting content can be downloaded in CSV or JSON

[1] You find HUBII under the following URL: www.hubii.world.

format. The diarization processing pipeline is available under www.hubii.world/processing/diarization-01.

3 Future Research

In this paper, we present the concept of the HUBII platform. HUBII provides access to biosignal processing, modelling, and consumption capabilities for human traits and states based on biosignals for researchers. With HUBII, we aim to contribute to NeuroIS research by leveraging different types of biosignals with three main aspects. First, HUBII makes biosignal processing, modelling, and consumption accessible to a larger group of researchers. Second, it makes biosignal insights not only accessible but also research much faster since researchers can build on the work of others. Third, it provides a standardized way for conducting research leveraging biosignals through its transparent nature. Finally, it allows researchers to develop HUBII-based applications such as user- and neuro-adaptive systems building on the capability of analyzing biosignals and predicting human traits and states.

We have planned two next steps for HUBII. Beforehand, we believe it is important to mention that HUBII has to be designed in accordance with ethical principles and data privacy rules. We plan to particularly keep this in mind when we develop HUBII together with the community based on a ethical conform DSR approach [10, 11]. Based on this assumption, we, first, want to further enhance the capabilities of HUBI and bring more pipelines in each of the layers on the platform. These pipelines should allow for processing more biosignals (e.g., audio, EDA, etc.) and derive features from them for further analysis. Second, HUBII shall become a community platform on which all members of the NeuroIS research community may contribute by allowing everybody to push stable pipelines to the platform. Our core objective is to let HUBII become a central portal for uploading biosignal datasets and models. This creates a living ecosystem and accessible value for future researchers.

References

1. Schultz, T., Amma, C., Heger, D., Putze, F., & Wand, M. (2013). Human-machine interfaces based on biosignals. AT-Automatisierungstechnik.
2. Seitz, J., & Maedche, A. (2022). Biosignal-based recognition of cognitive load: A systematic review of public datasets and classifiers. In *Information Systems and Neuroscience: NeuroIS Retreat 2022*.
3. Mai, B., & Kim, H. (2019). The relationships between emotional states and information processing strategies in is decision support—A NeuroIS approach. In *Information Systems and Neuroscience: NeuroIS Retreat 2019*. https://doi.org/10.1007/978-3-030-28144-1
4. Loos, P., Riedl, R., Müller-Putz, G. R., vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: Neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering, 2*, 395–401. https://doi.org/10.1007/s12599-010-0130-8

5. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems., 29*, 9–24. https://doi.org/10.1080/0960085X.2019.1708218
6. Jalilpour, S., & Müller-Putz, G. (2022). Toward passive BCI: Asynchronous decoding of neural responses to direction- and angle-specific perturbations during a simulated cockpit scenario. *Science and Reports, 12*, 6802. https://doi.org/10.1038/s41598-022-10906-5
7. He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems, 212*, 106622. https://doi.org/10.1016/j.knosys.2020.106622
8. Doyle, C., & Bailey, L. (2019). What can neurois learn from the replication crisis in psychological science? In *Information Systems and Neuroscience: NeuroIS Retreat 2019*.
9. Doyle, C., Luczak-Roesch, M., & Mittal, A. (2019). We need the open artefact: Design science as a pathway to open science in information systems research. In *Extending the Boundaries of Design Science Theory and Practice: 14th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2019*, Worcester, MA, USA, June 4–6, 2019, Proceedings 14.
10. Benke, I., Feine, J., Venable, J. R., & Maedche, A. (2020). On implementing ethical principles in design science research. In *THCI* (Vol. 12, pp. 206–227). https://doi.org/10.17705/1thci.00136.
11. Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*, 835–850. https://doi.org/10.1007/s10551-018-3921-3

NeuroIS at 15: What Were We Writing About?

Jacek Gwizdka

Abstract NeuroIS Retreat celebrates 15 years in 2023. To reflect on the topics presented at the retreats, we conducted two kinds of analyses of papers published in NeuroIS proceedings. First, an analysis of titles, abstracts and author-provided keywords (2015–2022); second, an analysis of full texts of all published NeuroIS proceedings (2011–2022). These analyses illustrate which topics remain in focus, how the topics have evolved over the past years, and which new topics come on the horizon. Our aim is to contribute to the field's retrospective and to support reflection.

Keywords NeuroIS · Text analysis · NLP · Retrospective · Topics · Methods · Tools

1 Introduction and Background

NeuroIS has emerged as an interdisciplinary field of research that integrates neuroscience and neurophysiological tools and theories with information systems (IS) research to gain deeper insights into the development, adoption, and impact of information and communication technologies. As the NeuroIS Retreat reaches its 15th anniversary in 2023, it's worthwhile to reflect on its history.

Inspired by the previous reviews (e.g., [1, 2]) and the survey of past NeuroIS Retreat attendees conducted to assess the current state of NeuroIS at the 10 year anniversary [3], we conducted a review and analysis of published NeuroIS proceedings using semi-automated methods and analyzed author keywords, titles, abstracts (2015–2022) and, separately, full-text of papers (2011–2022). In contrast to the previous reviews, we emphasize how the topics and methods have changed over time.

J. Gwizdka (✉)
School of Information, The University of Texas at Austin, Austin, USA
e-mail: NeuroIS2023@gwizdka.com

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_26

2 Method

Our text analysis (titles/abstracts/keywords and full text) approach rests on the assumption that theories, concepts, methods, and tools need to be expressed verbally to be presented and discussed. We used two sources of information, (1) titles, abstracts and author keywords from the proceedings published by Springer (2015–2022), and (2) full text extracted from PDF files of all published NeuroIS proceedings (2011–2022). We first describe our procedure for the full text processing.

In the full text NLP analysis, the frequency of word usage in the papers serves as an indicator of the prevalence of a specific topic, concept, method, or tool employed in research. The full text analysis was conducted using natural language processing (NLP) techniques in Python, the entire text of NeuroIS proceedings from 2011 to 2022 (as shown in Table 1) [4–14] was processed and analyzed. Full text of all papers (only abstracts were published during the years 2011–2013) was extracted from PDF files using `PyPDF2` library. Title pages, tables of contents, keynotes, and author indexes were skipped. Next, the text was processed using `nltk` and `spaCy`, the text was tokenized, lemmatized, and stop words were removed.

The list of keywords was processed by functions from `nltk` to obtain their frequency and the frequency of bigrams and trigrams. The resulting lists were then manually evaluated across all years, with consideration given to the definition of NeuroIS by Riedl et al. [15], NeuroIS fundamentals books [16, 17], and our general knowledge of the field. As a result of this process, a comprehensive list of 240 frequently occurring words and bigrams was compiled. Words and phrases were further aggregated based on their synonymy (see Table 3). For simplicity, we will refer to the aggregated words/phrases as *words*. For each individual and aggregated word in this list, its relative occurrence frequency (%) in each proceeding document was calculated. The use of relative frequency makes this analysis independent of the

Table 1 NeuroIS proceedings' characteristics 2011–2022

Year	No. of papers	Pages	Word count	Cleaned word count
2011	10	8	4,742	2,776
2012	22	24	16,271	9,322
2013	23	26	16,678	9,265
2014	15 + 15 posters	57	41,505	22,881
2015	29	221	58,261	33,939
2016	24	201	58,418	33,394
2017	24	214	60,791	35,481
2018	32	278	84,497	47,431
2019	40	359	110,275	62,962
2020	41	379	124,850	69,886
2021	27	262	83,683	45,875
2022	35	361	106,480	60,485

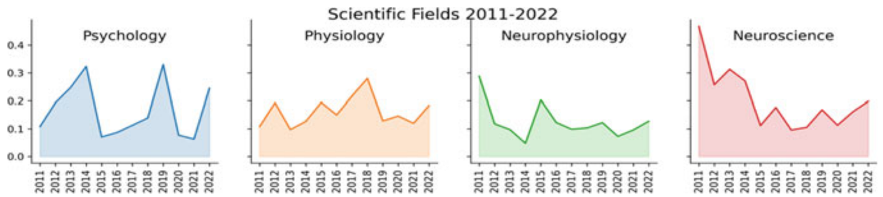


Fig. 1 Scientific fields mentioned in years 2011–22. In this and the following area charts, y-axis show the relative frequency (%) of a concept in each year (i.e., in each proceedings)

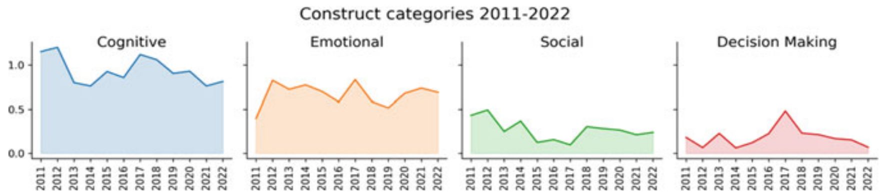


Fig. 2 The four main construct categories (2011–2022)

proceedings' length. However, less space in the shorter proceedings from 2011 to 2014 (in 2011–13 abstracts only, and in 2014 three-page-long papers) could affected word usage in these papers. The data set will be made available to other researchers. The relative frequencies of words/phrases across years 2011–2022 are shown in faceted area charts in figures: Figs. 1, 2, 4, 6, 8, 10, and 11. Although the variations in the frequency of words in proceedings (shown in these figures on y-axis) are meaningful indicators, our intention is not to make direct quantitative comparisons between words' relative frequencies, but rather to make qualitative comparisons and examine changes in each word's relative frequency of use over time.

In the analysis of titles, abstracts, and keywords, the author-provided keywords were manually aggregated based on their synonymy and utilized directly in the analysis. Author keywords assigned to NeuroIS papers 2015–2022 are summarized quantitatively in Table 2. We observe a high number of unique keywords. This makes us recommend using a controlled vocabulary for keywords, such as the ACM Computing Classification System, (https://dl.acm.org/ccs). Titles and abstracts were processed in a similar fashion as the full text, resulting in the inclusion of a few additional words that were added to the original author keywords—we will refer to this enlarged set of keywords as *all-keywords*. The synonymous keywords that referred to the same concept were combined, resulting in a reduced number of "processed" keywords (Table 2). The keywords were manually reviewed.

Due to the large number of papers (252) and associated keywords, we opted to analyze and visualize keywords that were assigned to at least two papers in a given year. Keywords which were assigned to only one paper in a year accounted for less than 5% of the total papers in that year, and were excluded from this analysis. Keywords included in the analysis were used to generate heatmaps (Figs. 3, 5, 9).

Table 2 NeuroIS author-provided keyword characteristics 2015–2022

Year	No. of papers	Keyword count	Unique keyword count	Proportion of unique keywords (%)	Processed keywords count	Proportion of unique processed keywords (%)
2015	29	133	120	90.23	94	70.68
2016	24	120	102	85.00	74	61.67
2017	24	126	114	90.48	90	71.43
2018	32	143	124	86.71	98	68.53
2019	40	196	171	87.24	131	66.84
2020	41	200	165	82.50	127	63.50
2021	27	133	119	89.47	96	72.18
2022	35	168	151	89.88	114	67.86

Additionally, these keywords were used to generate word clouds for years 2015–22 (Appendix Fig. 12). We believe that the heatmaps created from all-keywords and the area charts created from full-text analysis provide complimentary views of topics of interest in NeuroIS retreats.

3 Analysis

Our analysis aims to illustrate patterns emerging over time, as well as the variations in the prevalence of themes, topics, methods and tools.

Scientific Fields

Our analysis begins by examining the scientific fields which influence NeuroIS. It's hardly surprising that Psychology, Physiology, Neurophysiology and Neuroscience continue to be referenced in the NeuroIS papers (Fig. 1).

Construct Categories

Dimoka et al. [18] developed a list of 34 constructs of interest to IS research. The constructs were grouped into four high-level categories: cognitive, emotional, social and decision-making processes. We examined the same categories by using their respective keywords (including their synonyms) in the full-text analysis (Fig. 2).

Table 3 Selected synonyms applied in the text analysis

Concept	Synonyms
Cognitive (processes)	Cognitive (includes: cognitive process, cognitive processing, cognitive fit), cognition, perception, attention (including: attention, attentional control, attentional resource, concentration), perception, perceptual
Emotion; emotional (processes)	Affect, affective, affective processing, affective state, emotion, arousal, emotional
Cognitive load	Cognitive workload, cognitive load, cognitive-load, mental workload. mw, workload, mental effort, mental load, cognitive demand, cognitive effort, information overload, cognitive overload, memory load
Decision-making (processes)	Decision making, decision-making, decision inertia, decision maker, decision quality, making decision, make decision, decision process, decision uncertainty, decision making process, decision effort, decision support, decision phases, decision modeling, decision trees, decision delegation
Design science	Design oriented research, design theory, design research, design science, design science research, design artifact, design artefact, design process
Usability	Usability, usefulness, ease use, ease of use, user experience, UX
Consumer behavior	Consumer behavior, consumer decision making, online shopping, online payment, product choice, purchase decision, auction, neuromarketing
Information search	Information search, information avoidance, information seeking, information retrieval
Stopping behavior	Stopping behavior, stop search, stop seeking, stopping
Misinformation	Misinformation, misinform, disinform, false headline, false news, fake news
Bias	Bias, cognitive bias, bias susceptibility, processing bias
fMRI	fMRI, functional magnetic resonance
fNIRS	fNIR, fNIRS, near-infrared spectroscopy
EEG	EEG, electroencephalograph, electroencephalogram, consumer-grade EEG, research-grade EEG, mobile EEG
ERP	ERP, ERPs, event-related potential, event related potential, event-related brain potential, event related brain potential, perturbation-evoked potential
EFRP	EFRP, FRP, eye fixation related potential, eye-fixation related potential
Eye-tracking	Eye tracking, eye-tracking, eye tracker, eyetracking
Eye	Eye gaze, eye movement, eye blink, eye fixation, saccade
Pupil dilation	Pupil dilation, pupillary, pupillometry
Heart rate (HR); cardiovascular system measures	Heart rate, HRV, HR, ECG, blood pressure (BP)

(continued)

Table 3 (continued)

Concept	Synonyms
GSR	Skin conductance, EDA, electrodermal, electrodermal activity, galvanic skin response, GSR, SCR
Facial expression	Facial expression recognition, facial expression, facial emotion, facial activity

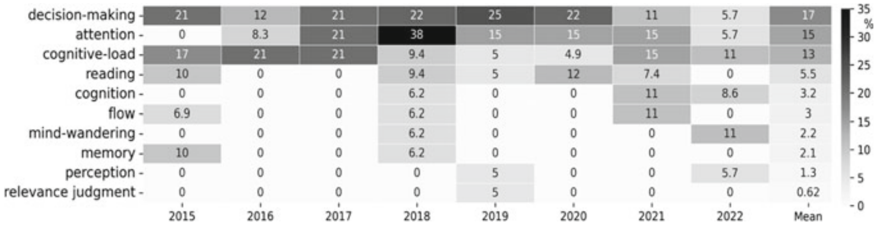


Fig. 3 Heatmap of topics related to cognitive processes created from all-keywords (NeuroIS papers 2015–2022). In all heatmaps, the cell values depict the percentage of papers associated with a keyword in a particular year in which the keyword was used. The rows in each heatmap are sorted on the average frequency of a keyword use per year (shown in the right-most column)

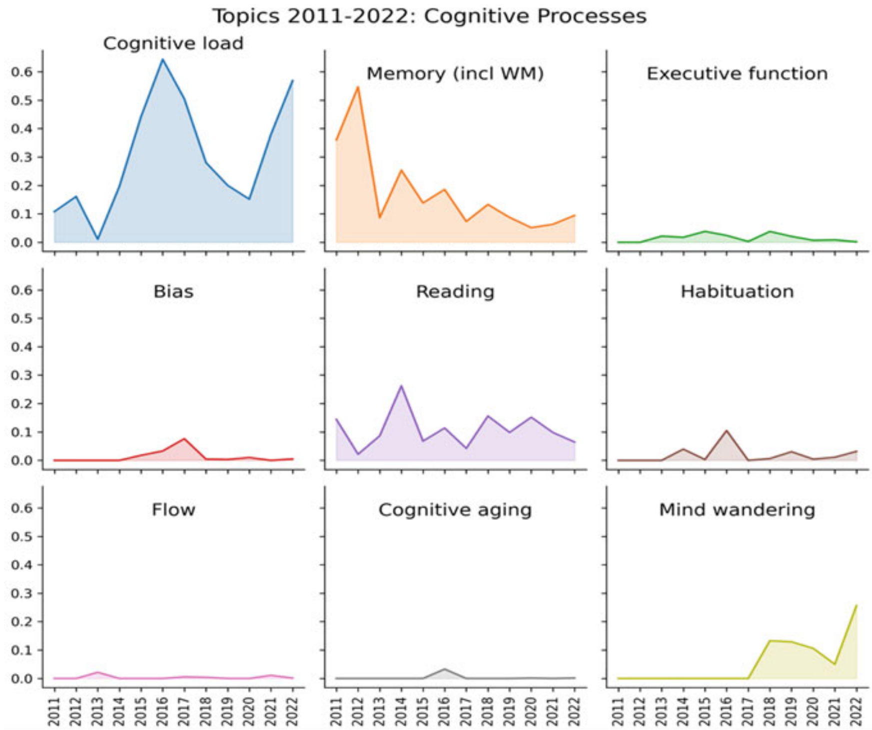


Fig. 4 Research topics related to cognitive processes across years 2011–2022

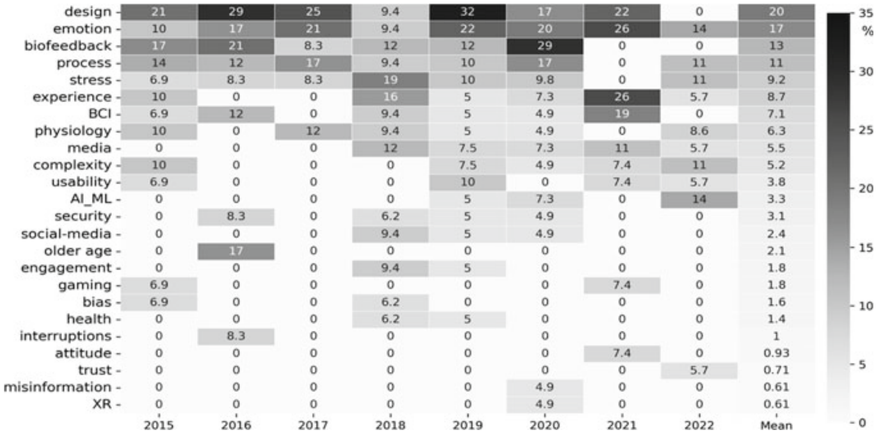


Fig. 5 Heatmap created from other topics in all-keywords (NeuroIS papers 2015–2022)

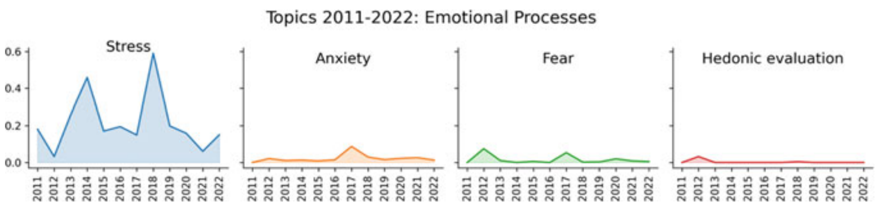


Fig. 6 Research topics related to emotional processes across year 2011–2022

Comparing Fig. 2 to 4.2 in Riedl and Léger [1] (where 43% papers were found to have a focus on cognitive, 39% on emotional, 14% on social and 3% on decision-making processes), we see similar proportions, except for decision-making processes, which became somewhat more prominent in NeruoIS research since 2014 (see also Fig. 3 in the next section). This indicates that the focus on high-level constructs described in 2011–2014 NeuroIS papers [1] remain roughly similar across 2011–2022.

Research Topics

Next, we analyze *all-keywords*, where each keyword is assigned to specific papers and thus, enables us to show proportion (%) of papers associated with that keyword in each year.

First, we discuss topics topics related to cognitive processes (Figs. 3 and 4). Topics related to decision-making, attention and cognitive load continue to be of strong interest in NeuroIS community. Decision-making-related papers frequently constituted 20–25% of the proceeding's papers, and overall, 17% of all papers (2015–22). This is somewhat in contrast with the recent survey results [3], where 10% of respondents indicated their current focus on decision-making and only 6% as future focus.

Cognitive-load (CL) was proposed as one of the opportunities in NeuroIS in the 2012 seminal article by Dimoka et al. [19]. Accordingly, we see a clear interest in CL across years. While CL-related papers were more frequent in earlier years (2015–2017: ~20%), more recently (2018–2022) there were fewer papers related to CL (5–15%). This observation is in line with the 2018 survey, where 25% respondents indicated CL as of past importance and only 5% said that CL needs more attention [3].

Although papers related to attention make up 15% of all papers, the interest in this topic seems to have peaked between 2017 and 2018, but has since shown a decline. On the other hand, mind wandering is a relatively new topic that emerged around 2018.

Full-text analysis (Fig. 4) generally shows similar patterns. A few topics not present in the heatmap, but present in the area chart, include executive function, habituation, and bias. Since 2015 we observe some, but still very small, interest in bias ("bias" includes cognitive bias, bias susceptibility, and processing bias).

We also examined other topics in the all-keywords and in the full-text analysis.

Papers related to design (design science/research) represent 20% of all papers. Biofeedback and process (process science) constitute just above 10% of all papers. The NeuroIS community has consistently demonstrated varying levels of interest in physiology-related topics (overall mean 6.3%).

Authors of the survey conducted among the NeuroIS attendees at the 10th anniversary [3] observed that: "emotional processes will likely be a key topic". We see a steady interest in emotional processes from the beginning of the NeuroIS retreat (Figs. 2 and 5). This interest seems to have always been there and so it is hard to draw a conclusion about emotions becoming a key interest.

Although stress can have physical and cognitive components, it is often considered to be an emotional process and thus it is shown as one of the emotional processes (Fig. 6). Stress is one of the areas with a strong interest. All-keyword analysis shows an average of 9.2% stress-related papers, with the peak of 19% in 2018 and no papers in 2021 (Fig. 5), while the full-text analysis (Fig. 6) shows somewhat weaker interest in stress in 2021–22. This finding may be partially reflecting the 2018 survey results, where over 40% respondents indicated that stress is of "past importance" [3].

From the full-text analysis (Fig. 7), there appears to be a decreasing interest in several areas, such as design science, and information search. Furthermore, the level of interest in technostress, usability as well in technology acceptance appears to be declining.

A few new topics of interest are emerging (Figs. 5 and 8). The interest in extended reality (XR), artificial intelligence (AI) / machine learning (ML) and misinformation has emerged, with a noticeable increase each year since 2017/2018. This is not surprising, given the recent developments in technology and society.

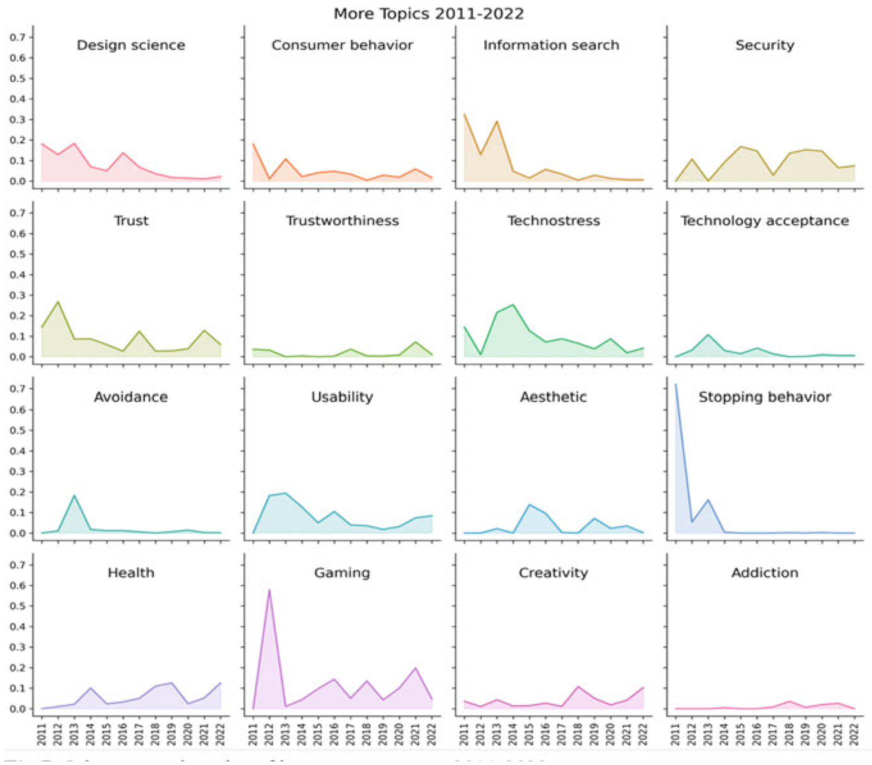


Fig. 7 Other research topics of interest across year 2011–2022

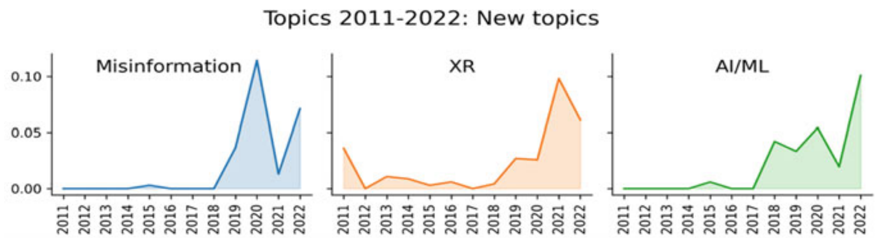


Fig. 8 New research topics of interest across years 2011–2022

Methods and Tools

We analyzed all-keywords and full-text to examine the methods and tools which are used in NeuroIS research. Our findings match to some extent findings from Riedl & Léger [1] (Fig. 4.3), Riedl et al. [2] and the results from the survey [3]. We note that [1] reviewed papers published at NeuroIS retreat 2011–2014, while [2] reviewed a much broader set of publication venues (55 journals and 13 conferences). The latter

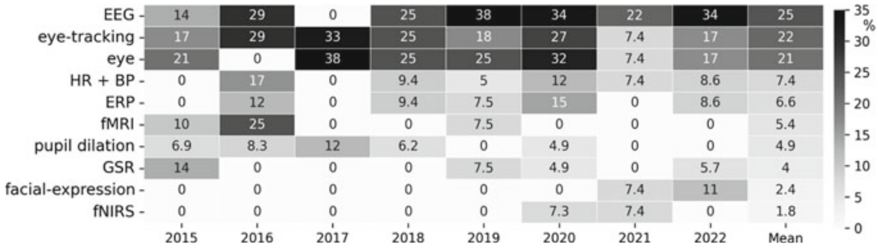


Fig. 9 Heatmap created from method and tool all-keywords (NeuroIS papers 2015–2022)

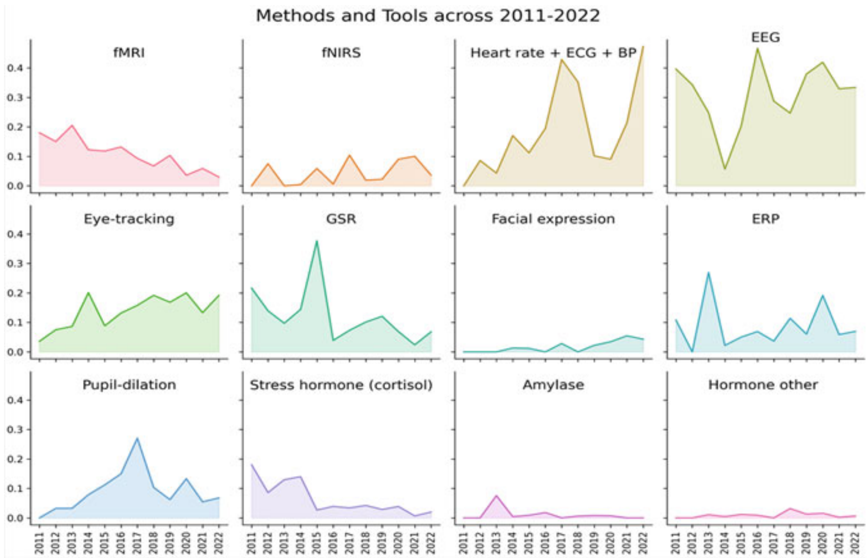


Fig. 10 Measurement methods and tools across years 2011–2022

may be the reason behind the differences between findings reported in [2] and in this paper.

EEG and eye-tracking are the two top methods/tools (all-keyword analysis shows 25 and 22% papers on the average respectively), and thus most likely they will remain of high relevance in the future. EEG continues to be one of the most commonly used tools, with the past reviews showing 25% in 2011–2014 [1] and 37% in 2008–2017 [2].

The aggregated keyword “eye” is closely related to eye-tracking, it includes eye-related measures (such as fixations, saccades, blinks, but not pupil dilation, which is considered here separately). Eye-tracking measures without pupil dilation are not

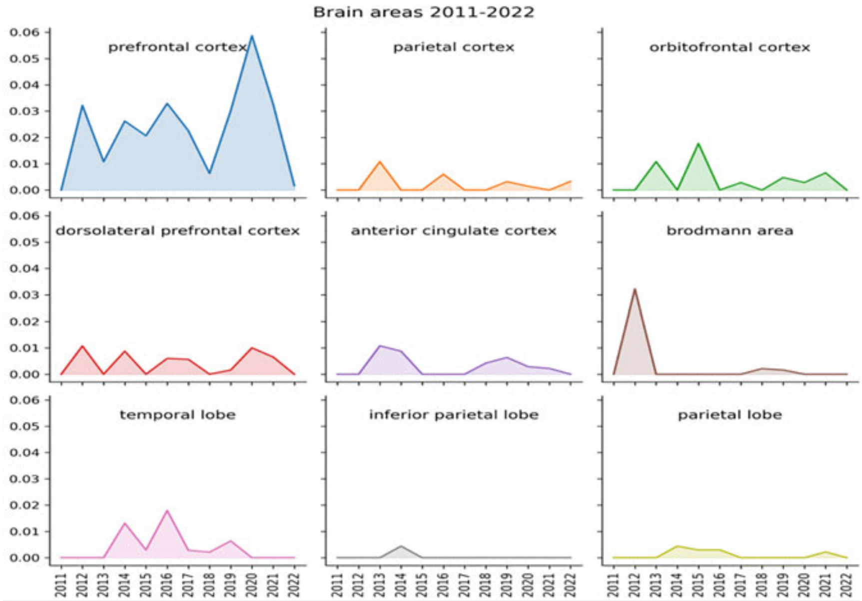


Fig. 11 Specific brain areas mentioned in NeuroIS proceedings 2011–2022

always considered as constituting NeuroIS (e.g., Riedl et al. [2]). We argue that they should be considered NeuroIS measures since the phenomena, such as eye fixation duration and gaze direction, are controlled by the nervous system, hence similarly as Riedl and Léger [1] we include them in Fig. 9, but, for clarity, we separate pupil dilation. Pupil dilation measurement was employed in most years (as seen from full-text analysis in Fig. 10). The average number of papers per year is 4.9% which is lower than 10% reported in Riedl et al. [2] for 2008–2017. The all-keyword analysis shows absence of pupil dilation in papers in 2019, 2021–2022 (Fig. 9).

The cardiovascular system measures (heart rate (HR) and electrocardiogram (ECG)) were mentioned every year starting 2012 (Fig. 10). The average proportion of papers using these measures is similar to findings reported in Riedl and Léger [1] at around 8%, but lower than 22% reported in Riedl et al. [2]. However, fMRI and GSR seem to show a much lower pick-up, namely 5.4 and 4%, respectively. This can be contrasted with 16% (fMRI) and 15% (GSR) in 2011–2014 NeuroIS papers [1] and 15% (fMRI) and 19% (GSR) in 2008–2018 [2]. The fMRI and GSR tendency observed by us in all-keyword analysis is confirmed by the trends shown in Fig. 10. Less frequent use of fMRI could, perhaps, be explained by its potentially lower acessibility and the cost [19]. GSR was mentioned almost every year, but there seems to be a trend of diminishing interest over time. We are unsure why the number of papers which use GSR is shrinking. Perhaps as more methods and tools become available, the average use of each becomes less frequent.

The interest in EMG suggested by the survey results [3] is not confirmed, as we did not find any mention of EMG. While the interest in various aspects of facial expression has been growing since 2017 (Fig. 10), it remains relatively small.

Brain Areas

We examined phrases (bigrams and trigrams) extracted from proceedings full-text for the names of brain areas. We identified nine areas shown in Fig. 11. These areas are not meant to be non-overlapping or non-mutually-exclusive. We found that prefrontal cortex continues to have a growing presence. Brodmann area was initially present in 2012 and then mentioned in 2018–2019. Other brain areas come occasionally into focus.

4 Conclusions

NeuroIS is a dynamic and exciting field that offers many opportunities for interdisciplinary collaboration between neuroscience and IS researchers and beyond. It has the potential to make significant contributions to our understanding of human behavior and decision-making in the digital age. The analyses presented in this paper illustrate how the topics of interest, methods and tools have evolved over the past 12 years, highlighting the increasing breadth of interests within the field. These findings are expected to stimulate reflection and promote further advancements in NeuroIS research.

Our approach has a number of limitations. First, the analysis based on titles, abstracts and keywords is mainly dependent on the keywords provided by papers' authors. This emphasizes the importance of selecting good descriptors by authors. The large number of unique author keywords points us to recommend creating (or adopting) and using a controlled vocabulary. Keywords from controlled vocabulary could be combined with more specific keywords supplied by authors.

Second, our full-text analysis used proceedings' PDF files. The process of extracting text from digitally-born PDF documents may not be completely accurate, as errors can occur in extracting ligatures, such as "ff" and "fi". Additionally, there is the possibility of human bias, as the use of words in papers depends on the authors' writing styles, with some authors tending to repeat words more often than others. Additionally, the personal beliefs and knowledge of the authors of this paper could have potentially influenced the selection of keywords for the analysis, which could have led to the omission or misclassification of important concepts.

Appendix

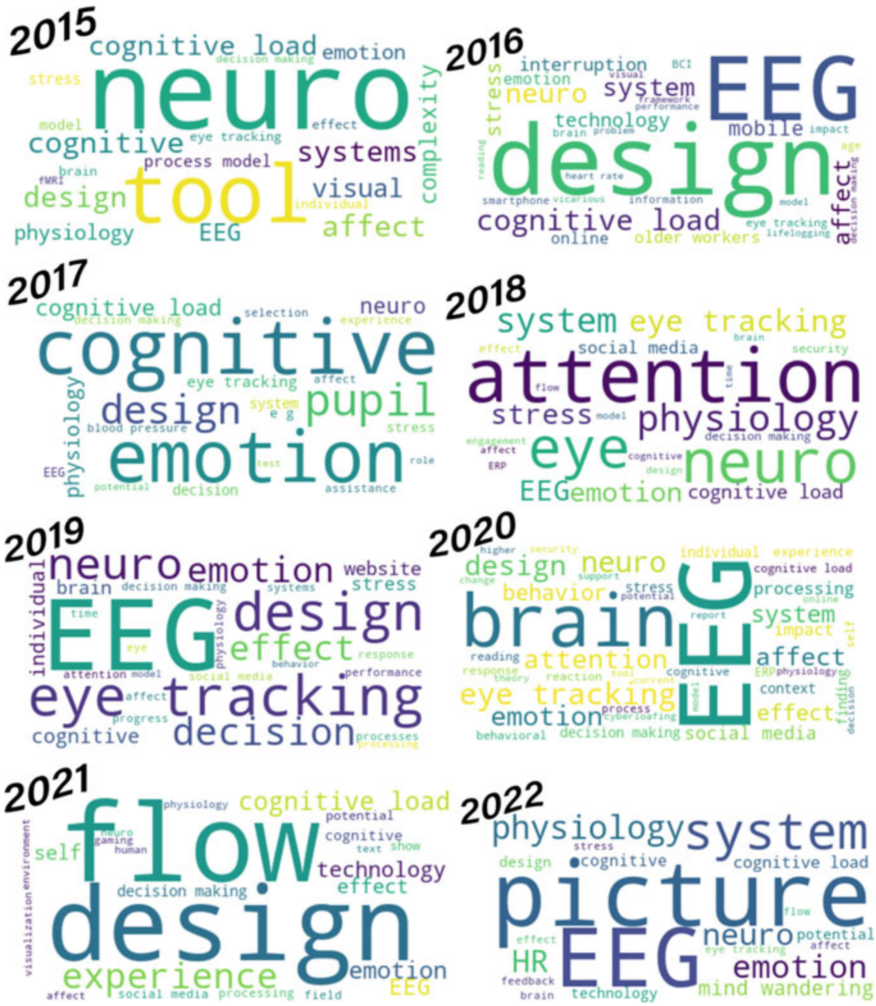


Fig. 12 Word clouds created from titles, abstracts & keywords (NeuroIS papers 2015–2022)

References

1. Riedl, R., & Léger, P.-M. (2016). Topics in NeuroIS and a taxonomy of neuroscience theories in NeuroIS. In *Fundamentals of NeuroIS* (pp. 73–98). Springer. https://doi.org/10.1007/978-3-662-45091-8_4
2. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D.: A decade of NeuroIS research: Progress, challenges, and future directions. ACM SIGMIS Database DATABASE. *Advanced Information Systems, 51*, 13–54 (2020). https://doi.org/10.1145/3410977.3410980
3. Fischer, T., Davis, F. D., Riedl, R. (2019). NeuroIS: A survey on the status of the field. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information systems and neuroscience* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-030-01087-4_1
4. Davis, F., Riedl, R., vom Brocke, J., Leger, P.-M., & Randolph, A. (2011). Proceedings Gmunden retreat on NeuroIS 2011.
5. Davis, F., Riedl, R., vom Brocke, J., Leger, P.-M., & Randolph, A. (2012). Proceedings Gmunden retreat on NeuroIS 2012
6. Davis, F., Riedl, R., vom Brocke, J., Leger, P.-M., & Randolph, A. (2013). Proceedings Gmunden retreat on NeuroIS 2013.
7. Davis, F., Riedl, R., vom Brocke, J., Leger, P.-M., & Randolph, A. (2014). Proceedings Gmunden retreat on NeuroIS 2014.
8. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., & Randolph, A. B. (eds.). (2015). *Information systems and neuroscience: Gmunden retreat on NeuroIS 2015*. Springer International Publishing. https://doi.org/10.1007/978-3-319-18702-0
9. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., & Randolph, A. B. (eds.). (2017). *Information systems and neuroscience: Gmunden retreat on NeuroIS 2016*. Springer International Publishing. https://doi.org/10.1007/978-3-319-41402-7
10. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., & Randolph, A. B. (eds.). (2018). *Information systems and neuroscience: Gmunden retreat on NeuroIS 2017*. Springer International Publishing. https://doi.org/10.1007/978-3-319-67431-5
11. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., & Randolph, A. B. (eds.). (2018). *Information systems and neuroscience: NeuroIS retreat 2018*. Springer International Publishing. https://doi.org/10.1007/978-3-030-01087-4
12. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A., & Fischer, T. (eds.). (2019). *Information systems and neuroscience: NeuroIS retreat 2019*. Springer International Publishing. https://doi.org/10.1007/978-3-030-28144-1
13. Davis F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Müller-Putz, G. (eds.). (2021). *Information systems and neuroscience: NeuroIS retreat 2021*. Springer International Publishing. https://doi.org/10.1007/978-3-030-88900-5
14. Davis, F. D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A. B., & Müller-Putz, G. R. (eds.). (2022) *Information systems and neuroscience: NeuroIS retreat 2022*. Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9
15. Riedl, R., Banker, R., Benbasat, I., Davis, F., Dennis, A., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P., Straub, D., vom Brocke, J., & Weber, B. (2010). On the foundations of NeuroIS: Reflections on the Gmunden retreat 2009. *Communications of the Association for Information Systems, 27*.
16. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS*. Springer. https://doi.org/10.1007/978-3-662-45091-8
17. Riedl, R., Davis, F. D., Banker, R., H. Kenning, P H. (2017). *Neuroscience in information systems research.* Springer International Publishing. https://doi.org/10.1007/978-3-319-48755-7
18. Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research, 22*, 687–702. https://doi.org/10.1287/isre.1100.0284

19. Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P. H., Pavlou, P. A., Müller-Putz, G., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in is research: Developing a research agenda for neurois. *MIS Quarterly, 36*, 679-A19.

Combating Videoconference Fatigue: A Pilot Study on the Effects of Video Layouts

Burak Öçlü, René Riedl, Eoin Whelan, and Thomas Acton

Abstract The COVID-19 pandemic has forced people to work, study, and communicate remotely, leading to a surge in the popularity of videoconferencing. However, prolonged and/or inappropriate use of videoconferencing tools may result in videoconferencing fatigue (VCF). This work-in-progress paper proposes an experimental research design which aims to determine how different video layouts are associated with VCF. A pilot study explored the impact of different video layouts on VCF, stress, engagement, and multitasking during a videoconferencing session. The pilot study, despite its limitations, offers first insights into the effects of video layouts on the well-being of participants and can provide a foundation for future research.

Keywords Videoconference fatigue (VCF) · Zoom fatigue · Digital stress · Video layout · Distant education · Experimental design · Technostress

B. Öçlü (✉) · E. Whelan · T. Acton
University of Galway, Galway, Ireland
e-mail: burak.oclu@universityofgalway.ie; b.oclu1@nuigalway.ie

E. Whelan
e-mail: eoin.whelan@universityofgalway.ie

T. Acton
e-mail: thomas.acton@universityofgalway.ie

R. Riedl
University of Applied Sciences Upper Austria, Steyr, Austria

Johannes Kepler University Linz, Linz, Austria

R. Riedl
e-mail: rene.riedl@fh-steyr.at; rene.riedl@jku.at

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_27

1 Introduction

Due to the global lockdown and social distancing policies initiated during the COVID-19 pandemic, videoconferencing tools were massively adopted to enable people continue to work, study, and interact socially. Some of the well-known videoconferencing solutions that are widely adopted include Zoom, Microsoft Teams, Google Meet, Skype, and Cisco Webex.

Videoconferencing solutions are particularly important in education because they allow students to learn from a distance. During the pandemic, universities all over the world had to switch to remote learning to maintain their teaching activities. Students could connect with their teachers and classmates through videoconferencing platforms. Even though videoconferencing offers advantages such as flexibility in time and place, some drawbacks have recently been noted. Users interact through a screen which creates differences compared to face-to-face meetings. Missing body language and facial expressions, self-awareness (if the camera is turned on), and multitasking activities are some factors that cause unnatural communication during videoconferencing meetings [1–4]. The consequences of unnatural communication are associated with increased fatigue and stress [5]. The videoconferencing fatigue (VCF) phenomenon also commonly referred to as "Zoom Fatigue"—which is characterized by feelings of exhaustion, depressive symptoms, and burnout [6]—has been linked to the prolonged and/or inappropriate use of videoconferencing tools. This condition is often accompanied by related symptoms such as anxiety, worry, burnout, discomfort, and stress, as well as physical symptoms like headaches [5].

Despite the lifting of social distancing restrictions, one recent study reports that 80% of students want to continue taking online classes after the pandemic [7]. Against this backdrop, it is important to understand the factors driving VCF so that the problem can be mitigated. With videoconferencing tools expected to continue to be used in education beyond the COVID-19 pandemic, research into its efficacy and impacts is warranted. The research question addressed by this study is: *Is there a significant difference in the level of fatigue experienced by videoconferencing participants using speaker view versus gallery view and how does this correlate with further variables such as stress, engagement, multitasking, and learning outcome?* Specifically, in this paper, we outline an experimental design, report the findings of a pilot study, and thereby establish a foundation for future research.

2 Problem Statement

Concentrating on both the screen and the presenter during a lecture can be mentally exhausting and may lead to fatigue. Anh et al. [8] found that the strength of the negative relationship between VCF and user satisfaction varied significantly between remote learners and remote workers. Remote learners reported reduced satisfaction as a result of VCF, compared to remote workers. According to Riedl [5], VCF can be caused by factors such as self-awareness, interaction with multiple faces that feel unnatural, and multitasking. However, there are few empirical studies which examine the relationship between videoconferencing layout options and VCF.

A videoconferencing setting that has been examined widely is the use of a webcam during meetings [1, 9–12]. However, while the video layout is one of the most used software settings [13], there is limited evidence to support its association with VCF. Recent studies indicate that individuals tend to prefer the gallery view over the speaker view (see Fig. 1) as their default video layout during videoconferencing.

It follows that the identification of an optimal video layout may aid in reducing VCF. For instance, using the speaker view layout during online classes can help students concentrate on the teacher and ignore other participants. However, it can also be mentally tiring as students have to maintain virtual eye contact with the instructor. Also, despite the fact that the gallery view layout can show all (or at least many) of the participants in the virtual room, it can lead to VCF as it is unnatural to have the feeling of being stared at by many people. Thus, the purpose of this study is to investigate how different video layouts affect VCF through empirical research (Fig. 2).

We hypothesize: *Participants using speaker view experience less fatigue than those using gallery view.* This hypothesis is based on the difference in cognitive load between the two layouts. The speaker view shows one large window that captures

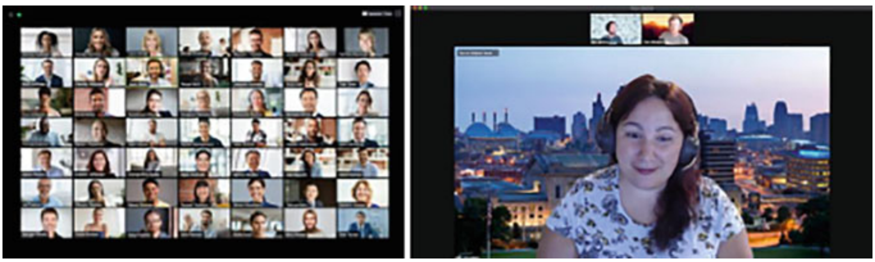


Fig. 1 Gallery view (left) and speaker view (right) layout based on the Zoom videoconferencing tool (*Source* [14, p. 6])

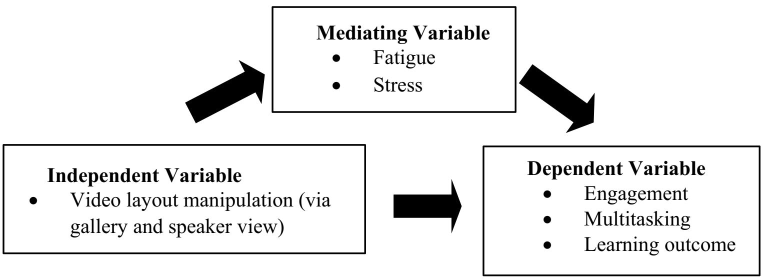


Fig. 2 Research model of the study

every-one's attention, while gallery view shows many equal-sized windows which can cause participants to divide their attention between multiple faces [2, 5]. Next, we discuss more related work as a foundation for our theorizing.

3 Related Work

Balogova and Brumby [13] explain why people prefer the gallery over the speaker view during videoconference sessions. They found that people choose the speaker view to concentrate on the speaker, but it becomes difficult to follow when there are multiple speakers. Abramova et al. [1] discovered that participants are likely to select the gallery view during crowded meetings, because it allows participants to recognize a larger number of people compared to the speaker view. Salim et al. [15] posit that people tend to look at the participants from the same angle when using the gallery view, but this is challenging because each webcam is placed differently. What follows is that the structure of the videoconference session can influence people's video layout preferences. In remote learning, there is typically only one primary instructor in learning sessions, and the speaker view can increase students' attention.

Peper et al. [16] found that both video layouts require constant focus on a close object, which is not typical in regular face-to-face meetings. Research shows that constantly looking at a screen can lead to fatigue [4, 10, 17]. The speaker view can help maintain constant virtual eye contact during meetings. However, the gallery view provides more room to look around, allowing participants greater flexibility and possibly impacts fatigue development.

Dependency on the screen, reduced mobility and limited viewing range cause unnatural interaction. Videoconferencing participants may become distracted and engage in other activities during sessions, such as checking emails or engaging with their mobile phones. This kind of multitasking is one of the coping mechanisms adopted when unnatural interactions become stressful [16, 18]. Recent studies show that multitasking is positively correlated with VCF [3]. Riedl [5] proposed that natural communication will result in less fatigue. Li et al. [14] propose to investigate whether changing the unnatural interaction with multiple faces using different video layouts could reduce VCF. Following this logic, one may expect that students who set a video layout mimicking the actual class experience will be less fatigued.

4 Experiment Design

The most popular videoconferencing software is Zoom, with approximately 55% market share [19]. The pilot study was conducted using Zoom because it is widely used in distance education. There are two main video layouts in Zoom: speaker and gallery views (see Fig. 1). A between-subjects design was used where a similar number of participants tested each scenario (speaker and gallery views). Specifically, we delivered a between-groups learning session over Zoom where participants used either gallery or speaker layouts. The study employed two key questionnaires to appraise the pre-meeting and post-meeting states of the participants, with the aim of examining differences between the two video layouts in VCF. Table 1 summarizes the specific questionnaires.

A pre-session questionnaire, administered immediately before the videoconference session, gathered demographic information, general fatigue level, and anxiety levels. Demographic questions encompass individuals' age, gender, and previous participation frequency in videoconferencing meetings. The level of VCF experienced by participants after the videoconferencing session can be influenced by their pre-existing fatigue levels. To account for this, we have added the General Fatigue Scale to the pre-questionnaire to assess the participants' pre-existing fatigue levels and to gain a better understanding of how these levels may impact their VCF.

The Mirror Anxiety Scale aims to collect information about how people feel when they are confronted with their own appearance, how much they care about how they seem, and how this has affected their experiences using videoconferencing tools. The experimental findings potentially corroborate the existing body of literature, as they serve as an antecedent to the phenomenon of VCF. Meanwhile, it may be disclosed that video layouts have an impact on individuals with mirror anxiety.

The post-meeting questionnaire covered VCF, stress, engagement, and multitasking. The ZEF scale is widely used as a tool to measure VCF, encompassing different fatigue types such as general, visual, and emotional fatigue. It is possible to identify various fatigue types and stress levels by comparing different video layouts in the experiment. According to the presumed different cognitive load between the two video layouts, engagement may differ. OSE focuses especially on engagement in remote learning. Increased VCF and decreased engagement may reveal higher multitasking activities as coping behaviour. If the gallery view creates more VCF and less engagement, that may also increase multitasking activities during the videoconferencing session.

Table 1 Pre- and post-videoconference session questionnaires

Pre-meeting questionnaire	Post-meeting questionnaire
Demographic questions	Videoconference fatigue (VCF) (Zoom Exhaustion and Fatigue scale, ZEF) [20]
General fatigue scale [21]	Measurement items of self-reported stress measure [22]
Mirror anxiety scale [23]	Online Student Engagement scale (OSE) [24]
	Multitasking scale [25]
	Learning outcome (4 questions about meeting topic)

5 Pilot Study and Conclusion

The pilot study was conducted on the 8th of March 2023. An industry practitioner was invited to give a career advice talk to final-year Business Information Systems postgraduate and undergraduate students through Zoom. The guest speaker was aware of the experiment and agreed not to present any slides, as we want participant attention to be on the faces of the speaker and other participants. To ensure students were not fatigued from other classes, the videoconference session commenced at 9 am, the first session of the day for students. Other than complete a survey before and after the talk, students were not required to complete specific tasks during the session.

An invitation letter disseminated to students prior to the talk provided them with a brief overview of the experiment and its requirements. Prior to the start of the videoconferencing session, a consent form was administered, which is included in the pre-questionnaire to be completed by the participants. As an incentive for meeting all the requirements, the participants were offered the chance to receive one of eight gift vouchers valued at 20 euros through a random draw. 34 participants gave their consent and attended. However, the pre-questionnaire and post-questionnaire were completed by only 17 participants (10 female and 7 male), with 7 participating in gallery view and 10 in speaker view. The videoconferencing session took 45 min. During this period, participants listened to the guest speaker in a randomly assigned video layout. Each participant's video layout was randomly assigned by the pre-questionnaire.

Cognisant of the small sample size and its impact on robustness and inference of results and potential to extrapolate, we conducted some exploratory independent sample between-group t-tests to provide a tentative exposition of differences in mediating and dependent variables, intended to help refine our model and better position the study for a follow-on experiment with a larger sample size.

Interestingly, although participants were randomly allocated to one of the 2 groups, initial t-tests showed differences in mean initial fatigue levels, with those who would soon experience the session in gallery view showing a pre-session higher fatigue (M $= 2.48$, SD $= 0.21$) than those who would experience the speaker view (M $= 1.94$, SD $= 0.47$), with the corresponding t-test suggesting statistical significance ($t(15) = -2.78$, $p = 0.004$).

Cognisant of notable differences in pre-meeting fatigue levels between the groups, to explore whether these differences were impactful we conducted a mixed between-within-subjects analysis of variance to assess the impact of two different interventions (Speaker View and Gallery View) on participants' general pre-meeting fatigue and post-meeting VCF. We did not detect any significant interaction effect between video layout and fatigue across pre- and post-meeting measures, with Wilk's Lambda $=$ 0.99, $F(1, 15) = 0.11$, $p = 0.74$, and partial eta squared $= 0.007$. There was a significant main effect for fatigue, Wilk's Lambda $= 0.74$, $F(1, 15) = 5.05$, $p = 0.04$, with partial eta squared $= 0.25$, with both groups showing a reduction in fatigue scores across the two time periods (see Table 2). The main effect comparing the two types of intervention was not significant $F(1, 15) = 3.51$, $p = 0.81$, with partial eta squared = 0.19, with the analysis failing to detect any impact of video layout on fatigue.

The small sample size in the pilot study prevented statistical significance in our analyses, with our statistical tests serving mainly to explore whether our hypothesis had a sufficient basis for future investigation. As such, we conducted exploratory independent sample t-tests to compare VCF, stress, engagement, and multitasking time for participants. There were no significant differences ($t(15) = -0.952$, $p = 0.356$) in VCF scores for gallery view ($M = 1.94$, $SD = 1.04$) and speaker view ($M = 1.54$, $SD = 0.67$), no significant differences ($t(15) = -1.13$, $p = 0.272$) in stress scores for gallery view ($M = 3.54$, $SD = 2.32$) and speaker view ($M = 2.48$, $SD = 1.53$), no significant differences ($t(15) = -1.66$, $p = 0.116$) in engagement scores for gallery view ($M = 4$, $SD = 0.78$) and speaker view ($M = 3.5$, $SD = 0.454$), no significant differences ($t(14) = -0.845$, $p = 0.412$) in multitasking time for gallery view ($M = 3.57$, $SD = 2.86$) and speaker view ($M = 2.66$, $SD = 1.32$).

Overall, the initial pre-session differences in fatigue may suggest that fatigue persists through the videoconference session, because, although the participant numbers fall short of minima for confidence in our analyses, when we focus solely on the means, gallery view had higher values than speaker view for post-session VCF, stress, and multitasking time. The study also suggests that initial fatigue may lead to increased multitasking and a lack of maintaining attention through the videoconference session. Even though using gallery view may make participants feel more fatigued because they are constantly seeing multiple video streams at once, it could still lead to higher engagement among participants. In other words, seeing everyone in the meeting could help people feel more connected and engaged, even if it takes more effort to do so. This is particularly interesting if future experiments can investigate it further with statistically robust samples, because it may suggest that gallery views can lead to better student engagement. Further research with a larger sample size is needed to further investigate our research model as shown in Fig. 2. The proposed experimental design could be significant in investigating the interactions between different video layouts and VCF for future NeuroIS studies. For a more comprehensive review, more research may be done using facial expression recognition, EEG, or eye-tracking technology (for an overview, please see [26–28]. After the final experiment is conducted, the plan for a follow-up study is to conduct an eye-tracking experiment to determine if different videoconferencing layouts are associated with higher cognitive load. Moreover, as the VFC literature is rapidly evolving, it is possible that new insights will result in the revision, advancement, or

Table 2 Fatigue levels for the speaker mode and gallery mode across two time periods

Time period	Speaker view			Gallery view		
	n	M	SD	n	M	SD
Pre-meeting (general fatigue)	10	1.94	0.47	7	2.48	0.21
Post-meeting (VCF)	10	1.54	0.67	7	1.94	1.04

refinement of our model. It will be rewarding to see what insights future research will reveal.

Our plan is to conduct a full-scale experiment, following a similar design discussed in this paper, with 100 participants in September 2023. Our pilot study revealed important learnings which need to be accounted for in the full experiment. Firstly, a number of students joined the session through their smartphones. The connection bandwidth for these students was limited forcing them to switch off their cameras. As a result, in the gallery view, other students would not have seen these participants. In the meantime, the size of the screen can impact facial recognition, with varying effects depending on the device used. For instance, when using Zoom's gallery view on a mobile phone, faces may appear significantly smaller than they do on a PC or Laptop. Additionally, smartphones often require stabilization when the camera is in use. These discrepancies possess the potential to compromise the consistency of the experiment. In the full experiment, participants will be required to join on a PC or laptop and keep their cameras on. Secondly, it is crucial to emphasize the importance of filling out the pre-meeting questionnaire promptly to avoid confusion during the experiment due to late attendance. The invitation letter may serve as a suitable platform to clearly communicate this prerequisite. Furthermore, based on our pilot study, we observed that there was a disparity between the number of attendees and the completed questionnaires. Thus, it would be advisable to compare the number of surveys with the number of attendees before the meeting commences to ensure consistency. In case of significant differences, additional requests may be necessary to minimize any potential discrepancies.

References

1. Abramova, O., Gladkaya, M., & Krasnova, H. (2021). An unusual encounter with oneself: Exploring the impact of self-view on online meeting outcome. In *Proceedings of the International Conference on Information Systems (ICIS)* (pp. 1–17).
2. Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*, 1–6.
3. Bonanomi, A., Facchin, F., Barello, S., & Vilani, D. (2021). Prevalence and health correlates of online fatigue: A cross-sectional study on the Italian academic community during the COVID-19 pandemic. *PLoS ONE, 16*, 1–19.
4. Kushner, B. J. (2021). Eccentric gaze as a possible cause of "Zoom fatigue." *Journal of Binocular Vision and Ocular Motility, 71*, 175–180.
5. Riedl, R. (2022). On the stress potential of videoconferencing: Definition and root causes of Zoom fatigue. *Electronic Markets, 32*, 153–177.
6. Montag, C., Rozgonjuk, D., Riedl, R., & Sindermann, C. (2022). On the associations between videoconference fatigue, burnout and depression including personality associations. *Journal of Affective Disorders Reports, 10*, 1–5.
7. Zheng, M., Bender, D., & Lyon, C. (2021). Online learning during COVID-19 produced equivalent or better student course performance as compared with pre-pandemic: Empirical evidence from a school-wide comparative study. *BMC Medical Education, 21*, 1–11.
8. Anh, L. E. T., Whelan, E., & Umair, A. (2022). 'You're still on mute'. A study of video conferencing fatigue during the COVID-19 pandemic from a technostress perspective. *Behaviour & Information Technology,* 1–15.

9. Bennett, A. A., Campion, E. D., Keeler, K. R., & Keener, S. K. (2021). Videoconference fatigue? Exploring changes in fatigue after videoconference meetings during COVID-19. *Journal of Applied Psychology, 106*, 330–344.
10. Kuhn, K. M. (2021). The constant mirror: Self-view and attitudes to virtual meetings. *Computers in Human Behavior, 128.*
11. Oducado, R. M. F., Fajardo, M. T. R., Parreño-Lachica, G. M., et al. (2021). Is videoconference "Zoom" fatigue real among nursing students? *Journal of Loss & Trauma,* 1–3.
12. Ratan, R., Miller, D. B., & Bailenson, J. N. (2022). Facial appearance dissatisfaction explains differences in Zoom fatigue. *Cyberpsychology, Behavior, and Social Networking, 25*, 124–129.
13. Balogova, K., & Brumby, D. (2022). How do you Zoom?: A survey study of how users configure video-conference tools for online meetings. In *2022 Symposium on Human-Computer Interaction for Work* (pp. 1–7).
14. Li, J., Lamers, M. H., & Riedl, R. (2022). Fewer faces displayed simultaneously, less videoconference fatigue in distance learning? An experimental study. In *Proceedings of the International Conference on Information Systems (ICIS).*
15. Salim, J., Tandy, S., Arnindita, J. N., et al. (2022). Zoom fatigue and its risk factors in online learning during the COVID-19 pandemic. *Medical Journal of Indonesia, 31*, 13–19.
16. Peper, E., Wilson, V., Martin, M., et al. (2021). Avoid Zoom fatigue, be present and learn. *NeuroRegulation, 8*, 47–56.
17. Woods, M. (2021). *Technology impacts on employee engagement during covid-19. Research and innovation forum.*
18. Kirshenbaum, N., Davidson, K., Harden, J., et al. (2021). Traces of time through space. In *Proceedings of the ACM on Human-Computer Interaction* (pp. 1–20).
19. Vailshery, L. S. (2022). Market share of videoconferencing software worldwide in 2022, by program. In *Statista.*
20. Fauville, G., Luo, M., Queiroz, A. C. M., et al. (2021). Zoom exhaustion & fatigue scale. *Computers in Human Behavior Reports, 4.*
21. Chalder, T., Berelowitz, G., Pavlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research, 37*, 147–153.
22. Tams, S., Thatcher, J., Hill, K., et al. (2014). NeuroIS—Alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems, 15*, 723–753.
23. Fauville, G., Luo, M., Queiroz, A. C. M., Lee, A., Bailenson, J. N., & Hancock, J. (2023). Video-conferencing usage dynamics and nonverbal mechanisms exacerbate Zoom Fatigue, particularly for women. *Computers in Human Behavior Reports, 10*, 1–14.
24. Dixson, M. D. (2015). Measuring student engagement in the online course: The online student engagement scale (OSE). *Online Learning, 19.*
25. Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior, 106*.
26. Riedl, R., & Léger, P.-M. (2016). *Fundamentals of NeuroIS—Information systems and the brain.* Springer.
27. Müller-Putz, G., Riedl, R., & Wriessnegger, S. (2015). Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems, 37*, 911–948.
28. Dimoka, A., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A. R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G. R., Pavlou, P. A., Riedl, R., vom Brocke, J., & Weber, B. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS Quarterly, 36*(3), 679–702.

Gaining Physiological Insight into Satisfaction with XAI Explanations: A Call for Research

Thomas Fischer, Stefan Faltermaier, Dominik Stoffels, and Marina Fiedler

Abstract The staggering performance of prediction models based on machine learning (ML) algorithms has led to a boom in research interest in their application, but also led to the question how black box algorithms arrive at their results. To open the black box, explainable AI (XAI) approaches have been developed, which create approximated models that make the results of ML black box models more transparent to human proponents. This transparency is important for a multitude of reasons (e.g., to trust automated decision making), but not all explanations are equally well received by human explainees. Thus far, explanation satisfaction is mainly measured through self-reports, and the application of neurophysiological measures in this specific context is widely lacking. We review the existing research and make suggestions for future research directions, calling for NeuroIS research into measurement approaches that could be applied in this domain.

Keywords Explainable AI · XAI · Explanation satisfaction · Evaluation

1 Introduction

In recent years, applications of machine learning (ML) have increasingly become popular, in part driven by the development of new machine learning techniques [1] and the increased availability of large datasets [2]. Although algorithms that can be

T. Fischer (✉) · S. Faltermaier · D. Stoffels · M. Fiedler
University of Passau, Passau, Germany
e-mail: thomas.fischer@uni-passau.de

S. Faltermaier
e-mail: stefan.faltermaier@uni-passau.de

D. Stoffels
e-mail: stoffe11@ads.uni-passau.de

M. Fiedler
e-mail: marina.fiedler@uni-passau.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_28

considered early applications of ML have been in use for decades [3], so-called black box algorithms have proliferated more recently. The ability of applications that use such algorithms to outclass humans in certain tasks has also reached widespread interest (e.g., in the case of the game Go [4]) and even led to some professions being deemed at risk of getting replaced by machines in the near future (e.g., [5]). This accuracy comes at a price though, as these algorithms can be opaque, which does not allow for ready interpretation of their inner workings by humans [6]. This downside is addressed by explainable AI (XAI) approaches that attempt to open the black box and provide us with an explanation for the decision process of black box algorithms. While being able to understand how an algorithm arrives at a specific result is certainly a nice feature to have, it has also become a mandatory element of decision-making based on algorithms as part of the EU's General Data Protection Regulation (GDPR).

This circumstance has undoubtedly created an additional drive for the development of XAI approaches, and it also brings the question to the forefront when an explanation provided by such approaches is deemed satisfactory by its recipient (i.e., the explainee). Currently, assessing explanation satisfaction is mainly done through self-report measures (e.g., [7, 8]), which can be an arduous exercise for explainees. We therefore propose that integrating neurophysiological measures into the process of evaluating XAI explanations could be a fruitful endeavor and a potential research field that can also profit from existing NeuroIS research into the cognitive processes involved in information retrieval. To substantiate this idea, we first provide more information on XAI approaches and the characteristics that individuals look for when evaluating explanations. We then review existing measures of explanation satisfaction and conduct a systematic literature review to check for neurophysiological insights into the evaluation of XAI explanations. We then provide suggestions for future research in this area and discuss the value for XAI development and beyond.

2 Research Background

ML Black Boxes and XAI. According to Mitchell [9] "…machine learning is concerned with the question of how to construct computer programs that automatically improve with experience" (p. XV). As the name suggests, approaches in this domain are concerned with learning from data and they apply techniques that identify robust and reliable patterns [6] in order to make predictions (e.g., to detect objects in images, to analyze texts, or even to solve puzzles [10]). Based on the level of *explainability* that is inherent in ML algorithms, we can classify some of them as black box algorithms (e.g., Support Vector Machines, Ensemble algorithms, or Neural Networks) [11, 12]. Algorithms in this group have frequently been shown to outperform humans in otherwise tedious classification tasks.[1] In addition

[1] Refer to https://www.eff.org/de/ai/metrics for an overview of areas in which machine-based classifications already outclass human accuracy.

to automating tasks, contemporary ML models have also been shown to be able to generate human-like performance (e.g., with increased public interest in the context of ChatGPT [13]), which may change many areas of society and in particular the way we organize work (e.g., [14]). A black box can become problematic though, if we need to know why (or why not) the algorithm came up with a specific result and what indicators and relationships amongst them led to the result. For example, Lebovitz et al. [15] report on the case of medical experts using machine-based classifications of visual data to augment their decision-making. Integrating the machine-based results into their work failed in several instances due to frequent disagreements with the results of the machine, which ultimately led to these additional results being deemed irrelevant as the experts could not trace back how the system had come up with its results.

To enhance the usefulness of such applications it is important to be able to open the black box, in particular due to these types of algorithms being able to construct their models autonomously and change them with new data points over time (i.e., they have learning capabilities [12]). Being able to understand why an ML algorithm arrived at its results can help us to improve the robustness of employed models (e.g., across different ML approaches or across data sets [16]) or find potential biases in the training the data that have been used to establish a ground truth (e.g., to reveal discriminatory patterns [17]), amongst many potential desiderata [18]. The field of XAI strives to deliver these types of explanations [1], for example through post hoc analyses of black box ML algorithms that then create a more understandable line of reasoning for why certain decisions were made, with the extent of such explanations being available for a model increasing its overall *interpretability* [11, 12]. There is a vast array of potential explanations that can be given for the drivers behind the outputs of a ML algorithm, which is reflected in the currently available XAI approaches (e.g., local or global explanation models, model-agnostic or model-specific estimation approaches, and different presentation formats such as visualizations, textual descriptions, or rules; see [19] or [11]). Hence, to narrow down what should be offered to explainees, it is important to know which types of explanations can satisfy their information needs.

Explanation Process and Explanation Satisfaction. To summarize the role of XAI as a "translator" for ML-based systems, consider the process depicted in Fig. 1. As can be seen, XAI plays a crucial role, when an explainee develops an information need based on the results that a ML-based system has produced (e.g., in the case of image classification, the question may arise why an image was classified in a certain way). XAI helps us in this regard, as it approximates the inner workings of the ML system and provides us with explanations (e.g., the main criterion that led to the classification, such as the dominating color of the depicted fruit).

In line with process models for explanations in the context of XAI presented by Hoffmann et al. [7] and Langer et al. [18], we can now expect a couple of different outcomes. In the case of an *unsatisfying explanation* (i.e., an unfulfilled information need), (i) the individual may consult the XAI interface (i.e., a system that implements interactive interrogation of the XAI model [20]) again to find further clues, but it can

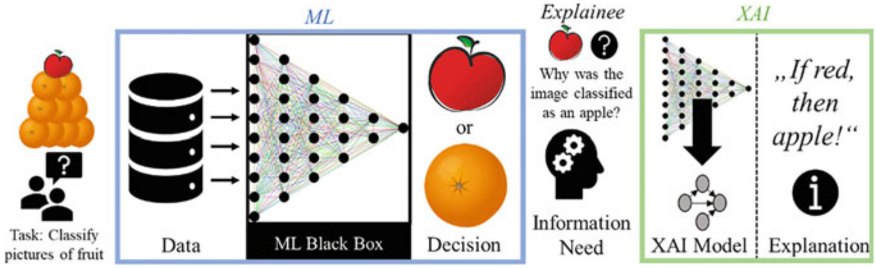


Fig. 1 Abstracted explanation process including explainer (machine) and explainee (human)

also occur that (ii) no further interactions happen despite an unsatisfying result (e.g., due to time constraints of the explainee [15]). In the case of a *satisfying explanation* (i.e., one that leads to an updated mental model in the individual on how the ML system got to its results), (iii) the individual may stop further interaction with the XAI interface, but there could also be circumstances that warrant further interaction. In the case that the individual is innately curious about the ML system and the domain it covers, (iv) further interrogations through the XAI interface may be started (e.g., further demands for explanations such as "When is an image classified as an orange?") [7]. In line with the further distinction between epistemic and structural satisfaction by Langer et al. [18], we can also expect that while an explanation can be sufficient to inform the explainee about the inner workings of a system (i.e., epistemic satisfaction), the explainee may still be dissatisfied with the system itself, as it may create its decisions in a way that is not acceptable (i.e., structural satisfaction). In this case, (v) the process could lead to an escalation targeted at the actual ML system and its controllers.

These different outcomes indicate the central role of explanation satisfaction in this process, and also highlight the need to distinguish the concept from related concepts such as curiosity and results satisfaction (i.e., structural satisfaction). For a working definition of explanation satisfaction, we refer to Hoffmann et al. [7] who define it as "…the degree to which users feel they understand the AI system or process being explained to them" (p. 4). Hence, explanation satisfaction is an effect that is achieved by an explanation, as opposed to a characteristic that is inherent to an explanation (i.e., such characteristics can be described as the "Goodness criteria" for an explanation [7]).

Based on this definition, we can also draw parallels with other forms of satisfaction that are popular in information systems (IS) research, including user satisfaction and job satisfaction, which can help us to form an initial idea of how explanation satisfaction measurement could be approached. For example, Bhattacherjee [21] discuss the formation of user satisfaction and the development based on the earlier concept of job satisfaction [22]. He highlights that satisfaction in the case of both constructs is represented by an affective state that results from a cognitive appraisal of the confirmation of expectations. The process involved is therefore largely the same, which we would also expect for other types of satisfaction such as explanation satisfaction, though the object that expectations are built towards (e.g., a digital technology in the case of user satisfaction, or a job in the case of job satisfaction) can be different. Hence, we would argue that experiences gained in the measurement of user satisfaction can also be widely transferred to the measurement of explanation satisfaction, though we have to be careful to distinguish between epistemic and structural satisfaction.

3 Measuring Explanation Satisfaction

Our argument for the need of additional measurement alternatives for explanation satisfaction is mostly of a practical nature. With the increasing prevalence of opaque systems, the need for explanations will likely also increase, and in some national contexts also already has a legal basis.[2] Though explanations are sometimes not offered for reasons that are independent of the specific technology used (e.g., corporate secrecy [24]), we argue that explanations are an important step for the further improvement of an AI system. As highlighted in Fig. 2, a successful explanation can act as a gateway that enables the explainee to consider potential improvements of blackbox systems. As pointed out before in the context of other types of satisfaction though, the expectations related to the object that is evaluated (i.e., the explanation in our case) can change [21]. In addition, which forms and formats of explanations will eventually lead to an understanding of a given subject in the explainee can be a highly individual matter [25]. In combination with the changing nature of expectations, this circumstance indicates that designing effective explanations will be an arduous task, which likely requires multiple design iterations and an uncertain level of individualization. Hence, in order to make this process more efficient and effective, a set of measurement approaches that allows for high levels of measurement accuracy at low levels of effort [26] for all stakeholders involved should be favored. In addition, physiological measures could be a complement to existing measurement approaches that helps to cover additional gaps in the interaction between users and AI systems (e.g., to enable users to give feedback on the quality of a system that goes beyond behavioral interactions [27]).

Non-physiological Measures. Based on the process depicted in Fig. 1 and its potential outcomes, we can assume that multiple ways to assess or approximate explanation satisfaction are imaginable. The first type of measures that could be used are self-report measures that directly assess the level of satisfaction with an explanation in

[2] The EU's GDPR [23] mandates that transparency is added to automated decision-making as is the case in ML-based classifications. In particular, § 15(1), § 22(3) and Recital 71(4) are relevant in this context as they lay out that individuals have the right to obtain information from an organization on the use of personal data about them (§ 15(1)b), its origin if it was not collected by the organization (§ 15(1)g) and whether automated decision-making is part its processing (§ 15(1)h). In case of automated processing, individuals also have the right to obtain an explanation on how the decision was reached, can challenge the decision, and have the right to obtain human intervention (§ 22(3) and Recital 71(4)).

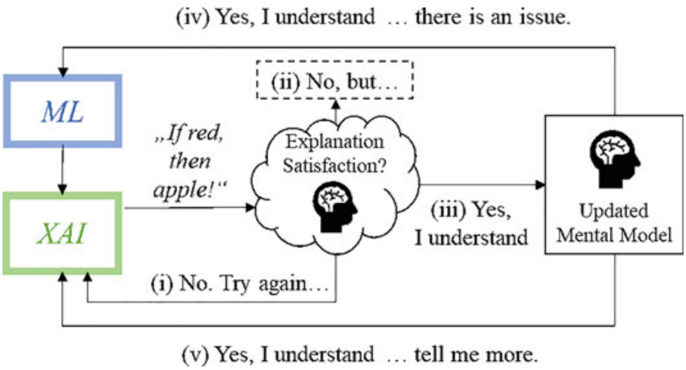


Fig. 2 Abstracted explanation process including explainer (machine) and explainee (human)

a standardized manner. In this group, we find instruments such as the Explanation Satisfaction Scale by Hoffmann et al. [7] (exemplary item: "This explanation of how the [software, algorithm, tool] works is satisfying") or the Systems Causability Scale by Holzinger et al. [8] (exemplary item: "I found that the data included all relevant known causal factors with sufficient precision and granularity"). Further types of self-reports that are common in the assessment of explanations are also an option, such as think-aloud tasks to examine whether the explanation solved the information need and improved the understanding of the ML system by the explainee [7]. A second group of measures that are related to this type of assessment of the resulting improvement in understanding are performance measures that test the knowledge level of the explainee before and after the explanation. A third group of measures could be behavioral measures that capture how the explainee interacts with the XAI interface.

All of these types of measures have inherent weaknesses though, which reduces their practical value as potential solitary measures of explanation satisfaction. For the mentioned self-report instruments, there are threats related to content validity or discriminant validity that can be mentioned as they also approximate explanation satisfaction through certain characteristics of explanation goodness (e.g., its level of detail) and its effects (e.g., on the achieved level of understanding of the problem). For both self-report and performance measures, their intrusiveness (i.e., how they interfere with a task [26]) also limits their usefulness in field research. For behavioral measures, we would require additional information on the cognitive processes of an explainee to determine whether an explanation was satisfying, as additional interrogations of the XAI interface or seizing interactions with the XAI interface can have different reasons aside from actual explanation satisfaction (e.g., curiosity or time constraints). These weaknesses point to the need for data triangulation and also suggest that it is worthwhile to pursue additional insights into the formation of explanation satisfaction.

Physiological Measures. To identify potential physiological measures that could be useful in this context, we conducted a structured literature review, which involved keywords related to the concept of explanation satisfaction [7, 20] and keywords related to methods commonly used in NeuroIS research.[3] Checking title and abstract of the resulting articles led to the identification of only two relevant articles. In the article by Polley et al. [28], eyetracking was used to gauge the attention (assessed through fixations) that individuals pay to explanations provided for the search results of a ML system that recommends books. In the study by Karran et al. [29], pupil dilation was used as an indicator of cognitive load in response to variations of characteristics of visualizations of XAI explanations. Though these studies are in the context of XAI explanations and use NeuroIS methods, neither of them is interested in explanation satisfaction specifically. We therefore extended our search process in an exploratory way (i.e., search in Google Scholar using variations of "satisfaction" with "information need" and NeuroIS terms) to identify studies that investigated the neural correlates of explanation satisfaction more specifically. This strategy proved to be more fruitful and resulted in the identification of the studies by Paisalnan and colleagues [30, 31].

Building on previous work that highlighted the distinct temporal progression in information retrieval tasks (i.e., information need realization, relevance judgment, information need satisfaction [32–34]), the authors conducted a series of fMRI studies to identify the neural correlates of information need satisfaction. Particularly relevant is the more recently published study in which the authors explicitly compared information need satisfaction to a baseline as well as the circumstance when obtained information does not satisfy an information need. They found that areas related to cognitive processes (i.e., the inferior frontal gyrus and the superior parietal lobe, which are involved in working memory and attentional control) and affective processes (i.e., the insula, which is involved in affective and evaluative processes) are involved in explanation satisfaction. This finding also indicates that when physiological measures are applied in the field (e.g., eyetracking or other measures that can be used outside of a laboratory setting), several sources of data have to be triangulated as multiple processes are involved in the formation of explanation satisfaction.

Though we find that there is a scarcity of research into the physiological measurement of explanation satisfaction, we find that the initial evidence found as part of the conducted review indicates that constructs and related measurement approaches that are also popular in NeuroIS research (i.e., cognitive load, attention, affect [35]) could be integrated into the further research on this topic. Though this finding also diminishes the possibility that one specific physiological indicator could be sufficient to detect the point when explainees are satisfied with an explanation, it does not rule out

[3] Scopus (01/17/2023): TITLE-ABS-KEY("explanation satisfaction" OR "explanation goodness" OR "explainability" OR "causability") AND TITLE-ABS-KEY("brain" OR "physio*" OR "eye*" OR "skin*" OR "heart*")—307 hits.

Web of Science (01/17/2023): TS = ("Explanation Satisfaction" OR "Explanation Goodness" OR "Explainability" OR "Causability") AND TS = ("Brain" OR "Physio*" OR "Eye*" OR "Skin*" OR "Heart*")—186 hits.

the possibility that physiological measures can be an important addition to existing non-physiological measures for related tasks such as design and evaluation cycles of XAI explanations. Due to this possibility and with the goal in mind to improve practical work on explanation quality, we want to suggest potential physiological measures and a research agenda in the next section.

4 Directions for Future Research

While we find that there is initial evidence for the neural correlates of explanation satisfaction, there is research needed in particular to identify indicators that can be used to implement its measurement outside of laboratory settings. Building on the findings by Paisalnan et al. [30], we should consider a combination of cognitive processes (attentional control, working memory) and affective processes (affective state of satisfaction) when selecting measures for further investigation. For the involved cognitive processes, eyetracking techniques could prove a worthwhile measurement approach. Pupil dilation as a measure of cognitive load has already been applied to evaluate XAI explanations [29], and eye movement behavior can be an indicator of spatial attention and attentional control [36]. For the affective processes involved and in particular the identified role of the insula, autonomic activity could also be used as a substitute to brain imaging techniques for measurement in the field. Insula activation is indicative of interoception and therefore reflects changes in autonomic activity [37]. As interoception is also involved in the formation of emotions [37], measuring changes in autonomic activities that have been linked to affect such as skin conductance or the startle reflex [38] could therefore prove valuable for this purpose. Combinations of measures in this way are viable as they have already been implemented in related fields such as the measurement of search satisfaction [39].

These measures in combination could help us to pinpoint the particular state that explainees reach when they are satisfied with an explanation. Indications for the existence of such a state can be found in the research on insight, sometimes also referred to as the “aha” or “Eureka” moment. In their review on the neural correlates of insight, Sprugnoli et al. [40] show that the neural processes found to be involved in information need satisfaction [30] have also been shown to be involved in the formation of the specific moment when individuals find the solution to a problem. Although this specific state is mostly related to creative problem solving, it could also be worthwhile to explore whether it can also occur in the moment when explainees understand why ML system has acted in a particular way.

If we consider the choice alternatives presented in Fig. 2, we also could approach explanation satisfaction indirectly, in the case that we find that there is no specific state that can pinpoint explanation satisfaction itself. Through experiments that vary the characteristics of an explanation (i.e., the aforementioned goodness criteria [7]) and then observe the resulting behavior, we could identify when the explainee requires further explanations and when the point of sufficient understanding is approached (i.e., through exclusion of alternative explanations such as insufficient time, or simple curiosity, as highlight in Fig. 2). Either through the direct approach (i.e., measuring when the explainee is satisfied) or the indirect approach (i.e., measuring indicators that could highlight the potential alternatives to explanation satisfaction), we could then try to improve the quality of explanations. Controlling also for factors that affect satisfaction (e.g., experience with a specific system, as is the case in the context of user satisfaction [41]), we could try to identify explainee characteristics and subject characteristics (i.e., the subject matter that is explained to the explainee) that help us to tailor explanations to the needs of the explainee. The goal could then be to create practical guidance for the design of XAI explanations (e.g., the most important design considerations that should be kept in mind) and a measurement routine that allows for the fast-paced evaluation of the effect of design changes.

Even outside of more practical contributions as the ones we are advocating for, there is still a great need for research on the evaluation of XAI systems. Investigating the potential physiological measures for the evaluation of XAI systems now could come at an opportune moment, as though there is already a host of potential measures and related taxonomies (e.g., [42, 43]), XAI evaluation is still in its early stages [44] with a pivotal indicator being that XAI systems are barely evaluated by researchers in general [19]. Hence, adding physiological indicators as an unobtrusive addition to the measurement portfolio could help to foster more frequent XAI evaluation. In addition, investigating explanation satisfaction physiologically could path the way for the development of neuroadaptive systems [45] that tailor explanations to the highly individual preferences of explainees [25]. Finally, further investigations of explanation satisfaction can also be of theoretical value, as it could, for example help to identify the sweet spot where an explanation is sufficient for an explainee, while keeping close to the accuracy of the estimated model [6] or help to illuminate the complex relationship between reported explanation satisfaction and actual understanding of a topic as well as related curiosity [46].

We consider the NeuroIS community well equipped to contribute in this area, as the phenomenon falls into the center of topics that are of interest to the NeuroIS community [35], which has also seen a trend towards measurement in the field recently (e.g., [47]). Measurement in the field using measures that could combine insights into the attentional and affective processes involved in explanation satisfaction could prove particularly worthwhile and the NeuroIS community has the necessary expertise in this regard (e.g., with additional techniques such as off-the-shelf EEG devices to capture cognitive load [48], eyetracking to capture attention [49], and Facereader to capture affect [50], that could easily be integrated in the workplace) [51].

5 Conclusion

We highlighted the value that explanation satisfaction has in the interaction of an explainee with a ML-based system and its XAI interface, and also pointed to the importance of investigation physiological indicators of explanation satisfaction. In particular, we call for research that investigates the physiological indicators that could be indicative of explanation satisfaction, which can also be employed in field settings to path the way for systems that are eventually able to tailor their explanations based on an individual user's information needs.

References

1. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI-explainable artificial intelligence. *Science Robotics, 4*. https://doi.org/10.1126/scirobotics.aay7120
2. Müller, O., Junglas, I., Vom Brocke, J., & Debortoli, S. (2016). Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *European Journal of Information Systems, 25*, 289–302. https://doi.org/10.1057/ejis.2016.2
3. Confalonieri, R., Coba, L., Wagner, B., & Besold, T. R. (2021). A historical perspective of explainable artificial intelligence. *WIREs Data Mining and Knowledge Discovery, 11*. https://doi.org/10.1002/widm.1391
4. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*, 484–489. https://doi.org/10.1038/nature16961
5. Reardon, S. (2019). Rise of robot radiologists. *Nature, 576*, S54–S58. https://doi.org/10.1038/d41586-019-03847-z
6. Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science, 32*, 856–880. https://doi.org/10.1287/orsc.2020.1382
7. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for explainable AI: Challenges and prospects.*
8. Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The system causability scale (SCS): Comparing human and machine explanations. *Kunstliche Intelligenz, 34*, 193–198. https://doi.org/10.1007/s13218-020-00636-z
9. Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
10. Samek, W., & Müller, K.-R. (2019). Towards explainable artificial intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 5–22). Springer International Publishing.
11. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
12. Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*, 1433–1450.
13. van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*, 224–226. https://doi.org/10.1038/d41586-023-00288-7
14. Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science, 358*, 1530–1534. https://doi.org/10.1126/science.aap8062
15. Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science, 33*, 126–148. https://doi.org/10.1287/orsc.2021.1549
16. Stoffels, D., Faltermaier, S., Strunk, K. S., & Fiedler, M. (2022). Opening the black-box of AI: Challenging pattern robustness and improving theorizing through explainable AI methods. In R. Beck, S. Petter, & T. B. Jensen (Eds.), *Proceedings of the International Conference on Information Systems (ICIS) 2022.*

17. Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In K. Martin (Ed.), *Ethics of data and analytics. Concepts and cases. An Auerbach book* (pp. 296–299). CRC Press Taylor & Francis Group.
18. Langer, M., Oster, D., Speith, T., Hermanns, H., Kästner, L., Schmidt, E., Sesing, A., & Baum, K. (2021). What do we want from explainable artificial intelligence (XAI)?—A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence, 296*, 103473. https://doi.org/10.1016/j.artint.2021.103473
19. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
20. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery, 9*, e1312. https://doi.org/10.1002/widm.1312
21. Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly, 25*, 351–370. https://doi.org/10.2307/3250921
22. Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297–1343). Rand McNally College Pub. Co.
23. Regulation (EU) 2016/679 of the European parliament and of the council (2016).
24. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*, 205395171562251. https://doi.org/10.1177/2053951715622512
25. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007
26. Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems, 15,* i–xxxv.
27. Tarafdar, M., Page, X., & Marabelli, M. (2023). Algorithms as co-workers: Human algorithm role interactions in algorithmic work. *Information Systems Journal, 33*, 232–267. https://doi.org/10.1111/isj.12389
28. Polley, S., Koparde, R. R., Gowri, A. B., Perera, M., & Nuernberger, A. (2021). Towards trustworthiness in the context of explainable search. In F. Diaz, C. Shah, T. Suel, P. Castells, R. Jones, & T. Sakai (Eds.), *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2580–2584). ACM. https://doi.org/10.1145/3404835.3462799
29. Karran, A. J., Demazure, T., Hudon, A., Senecal, S., & Léger, P.-M. (2022). Designing for confidence: The impact of visualizing artificial intelligence decisions. *Frontiers in Neuroscience, 16*, 883385. https://doi.org/10.3389/fnins.2022.883385
30. Paisalnan, S., Moshfeghi, Y., & Pollick, F. (2022). Neural correlates of satisfaction of an information need. In G. Nicosia & P. M. Pardalos (Eds.), *Proceedings of the 2nd International Advanced Course & Symposium on Artificial Intelligence & Neuroscience.*
31. Paisalnan, S., Pollick, F., & Moshfeghi, Y. (2021). Towards understanding neuroscience of realisation of information need in light of relevance and satisfaction judgement. In G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, G. Jansen, P. M. Pardalos, G. Giuffrida, & R. Umeton (Eds.), *Proceedings of the 7th International Conference on Machine Learning, Optimization, and Data Science, LOD 2021.* Springer eBook Collection (pp. 41–56). Springer International Publishing; Imprint Springer.
32. Moshfeghi, Y., & Pollick, F. E. (2018). Search process as transitions between neural states. In P.-A. Champin, F. Gandon, M. Lalmas, & P. G. Ipeirotis (Eds.), *Proceedings of the 2018 World Wide Web Conference on World Wide Web—WWW '18* (pp. 1683–1692). ACM Press. https://doi.org/10.1145/3178876.3186080
33. Moshfeghi, Y., & Pollick, F. E. (2019). Neuropsychological model of the realization of information need. *Journal of the Association for Information Science and Technology, 70*, 954–967. https://doi.org/10.1002/asi.24242

34. Gwizdka, J., Hosseini, R., Cole, M., & Wang, S. (2017). Temporal dynamics of eye-tracking and EEG during reading and relevance decisions. *Journal of the Association for Information Science and Technology, 68*, 2299–2312. https://doi.org/10.1002/asi.23904
35. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *The Database for Advances in Information Systems, 51*, 13–54. https://doi.org/10.1145/3410977.3410980
36. Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325
37. Gasquoine, P. G. (2014). Contributions of the insula to cognition and emotion. *Neuropsychology Review, 24*, 77–87. https://doi.org/10.1007/s11065-014-9246-9
38. Koller, M., & Walla, P. (2012). Measuring affective information processing in information systems and consumer research—Introducing startle reflex modulation. In AIS (Ed.), *Proceedings of ICIS 2012.*
39. Wu, Y., Liu, Y., Tsai, Y.-H.R., & Yau, S.-T. (2019). Investigating the role of eye movements and physiological signals in search satisfaction prediction using geometric analysis. *Journal of the Association for Information Science and Technology, 70*, 981–999. https://doi.org/10.1002/asi.24240
40. Sprugnoli, G., Rossi, S., Emmendorfer, A., Rossi, A., Liew, S.-L., Tatti, E., Di Lorenzo, G., Pascual-Leone, A., & Santarnecchi, E. (2017). Neural correlates of Eureka moment. *Intelligence, 62*, 99–118. https://doi.org/10.1016/j.intell.2017.03.004
41. Adam Mahmood, M., Burn, J. M., Gemoets, L. A., & Jacquez, C. (2000). Variables affecting information technology end-user satisfaction. A meta-analysis of the empirical literature. *International Journal of Human-Computer Studies, 52,* 751–771. https://doi.org/10.1006/ijhc.1999.0353
42. Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning.*
43. Chromik, M., & Schuessler, M. (2020). A taxonomy for human subject evaluation of black-box explanations in XAI. In *Proceedings of the IUI Workshop on Explainable Smart Systems and Algorithmic Transparency in Emerging Technologies (ExSS-ATEC'20).*
44. Lopes, P., Silva, E., Braga, C., Oliveira, T., & Rosado, L. (2022). XAI systems evaluation: A review of human and computer-centred methods. *Applied Sciences, 12*, 9423. https://doi.org/10.3390/app12199423
45. Vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*, 9–24. https://doi.org/10.1080/0960085X.2019.1708218
46. Liquin, E. G., & Lombrozo, T. (2022). Motivated to learn: An account of explanatory satisfaction. *Cognitive Psychology, 132*, 101453. https://doi.org/10.1016/j.cogpsych.2021.101453
47. Greif-Winzrieth, A., Peukert, C., Toreini, P., & Adam, M. T. P. (2021). Exploring the potential of NeuroIS in the wild: Opportunities and challenges of home environments. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, A. B. Randolph, & G. Müller-Putz (Eds.), *Information systems and neuroscience*. Lecture notes in information systems and organisation (Vol. 52, pp. 38–46). Springer International Publishing. https://doi.org/10.1007/978-3-030-88900-5_5
48. Conrad, C. D., & Bliemel, M. (2016). Psychophysiological measures of cognitive absorption and cognitive load in E-learning applications. In AIS (Ed.), *Proceedings of ICIS 2016*.
49. Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001
50. Giroux, F., Léger, P.-M., Brieugne, D., Courtemanche, F., Bouvier, F., Chen, S.-L., Tazi, S., Rucco, E., Fredette, M., Coursaris, C., et al. (2021). Guidelines for collecting automatic facial expression detection data synchronized with a dynamic stimulus in remote moderated user tests. In M. Kurosu (Ed.), *Human-computer interaction. Theory, methods and tools*. Lecture notes in computer science (Vol. 12762, pp. 243–254). Springer International Publishing. https://doi.org/10.1007/978-3-030-78462-1_18

51. Fischer, T., Davis, F. D., & Riedl, R. (2018). NeuroIS: A survey on the status of the field. In F. D. Davis, R. Riedl, J. Vom Brocke, P.-M. Léger, & A. Randolph (Eds.), *Information systems and neuroscience: NeuroIS retreat 2018*. Lecture notes in information systems and organisation (pp. 1–10). Springer.

Examining the Learner's Cognitive Load in Response to Different Learning Material in High and Low Immersive Virtual Learning Environments—An Eye-Tracking Study

Jana Gonnermann-Müller and Malte Teichmann

Abstract Learning in virtual, immersive environments must be well-designed to foster learning instead of overwhelming and distracting the learner. So far, learning instructions based on cognitive load theory recommend keeping the learning instructions clean and simple to reduce the extraneous cognitive load of the learner to foster learning performance. The advantages of immersive learning, such as multiple options for realistic simulation, movement and feedback, raise questions about the tension between an increase of excitement and flow with highly realistic environments on the one hand and a reduction of cognitive load by developing clean and simple surroundings on the other hand. This study aims to gain insights into learners' cognitive responses during the learning process by continuously assessing cognitive load through eye-tracking. The experiment compares two distinct immersive learning environments and varying methods of content presentation.

Keywords Eye-tracking · Cognitive load · Virtual learning environments · Learning performance · Immersion · Virtual reality

1 Introduction

Virtual reality offers a wide range of possibilities for education and learning, especially for gaining practical (*hands-on)* knowledge, e.g. through context-based information presentation and immediate transfer and application of knowledge in real-world simulations [1–3]. Moving through a virtual space fosters learning motivation

J. Gonnermann-Müller (✉) · M. Teichmann
Chair of Business Informatics, Processes and Systems, Potsdam, Germany
e-mail: jana.gonnermann@wi.uni-potsdam.de

Weizenbaum-Institute e. V, Berlin, Germany

M. Teichmann
e-mail: malte.teichmann@wi.uni-potsdam.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_29

by presenting content in a real scenario, increasing fun and arousing curiosity [4–6]. Learning in a virtual environments also facilitates the transfer and application of the content because it is linked to real-world activities and problems. However, to benefit from the given advantages, the virtual learning environments (VLE) must be designed accurately so that the learner benefits from immersive learning technologies. With a growing body of research, the awareness of the downsides of being in VLE, like cybersickness and mental fatigue, work to the surface [7, 8]. Both effects are related to adverse physical symptoms, like headaches, dizziness or exhaustion, caused by an overwhelming and stressful use of virtual reality. Properly designed VLE intend to foster learning and prevent negative physiological symptoms by providing supportive conditions, including instructional and didactical design, selection, and combination of learning material. One popular theory design guidelines for digital and online learning are so far based on is the cognitive load theory. Prominent examples are the *Cognitive Load Effects* [9], the *Cognitive Theory of Multimedia Learning (CTML)* [6] and the *four-component instructional design (4C/ID) model* [10]. The cognitive load theory is built on the assumption of a restricted working memory capacity that limits the processing of new information, which affects learning performance. For that reason, design guidelines recommend designing clean and simple learning instructions and environments, which decreases extraneous cognitive load, leaving enough resources for the learning [11]. However, since learning environments exceed new possibilities like a high virtual immersion, interaction and movement and intend to display highly realistic and stimulating VLE, this has questioned the previously existing assumptions of the cognitive load theory [9, 12]. Recent recommendations face contrasting aspirations to increase the learner's feeling of flow, which shows a positive relationship between flow and learning in computer environments [13].

Given that contrasting ambitions, authors demand to empirically investigate the connection between cognitive load theory and learning in VLE in more detail [9, 12]. Given this background, this experimental study examines the role of cognitive load (CL) during learning in VLE. It addresses the research questions:

RQ 1: Do a higher degree of immersion in the vle and a higher interactivity of the learning material increase the learner's cognitive load?

And furthermore:

RQ 2: Is there an association between immersion, interactivity of the learning material, and cognitive load with learning performance?

A two-step analysis is applied to answer the research questions. Furthermore, eye-tracking is applied to assess the CL to gain insights into changes in the learner's CL response over the whole learning phase, with a primary interest in changes in average pupil dilation and the variation of pupil dilation in response to learning in VLE [14–16].

Before addressing the relationship between VLE immersion, CL and learning performance, we first intend to give an overview of virtual reality for learning, CL and its measurement. The third chapter introduces the methodological procedure and the experimental approach. Finally, we provide an outlook on data analysis.

2 Theoretical Background and Hypothesis Development

Virtual Learning Environments

With virtual reality, it is possible to simulate a work environment in a virtual space using 3D models and simulations [7]. By linking abstract content, like hands-on knowledge, to 3D animations, knowledge gets contextualised, resulting in better learning outcomes. The simulation of dangerous actions or complex procedures in VLE enables a safe space for learning without negative consequences. Besides depicting realistic content, virtual reality features allow the learner to move through a virtual space while learning self-determined [4–6]. VLE can vary in immersion from highly immersive environments to low immersive environments. The degree of immersion is based on a technical dimension that arises from hardware features. Virtual reality cages (360° virtual environments) or virtual reality via a head-mounted display (HMD) are defined as highly immersive environments. In contrast, desktop VR is defined as a low-immersive environment [17]. Besides this technical dimension, a more subjective part of immersion depicts how realistic and involved the learner experiences the virtual environment. A high degree of immersion can positively affect the learner's enjoyment and motivation to act within the environment [12], resulting in a feeling of flow. The flow theory, often mentioned regarding gaming and virtual spaces [13], describes a psychological state in which the person feels energised, focused, and enjoyable when performing an activity or task. However, poorly-designed environments can cause problems such as cybersickness, mental fatigue, and cognitive overload for the learner. Cybersickness [7] and mental fatigue [8] describe physiological reactions to virtual environments. Cybersickness describes physiological responses such as discomfort, dizziness, nausea, headaches, or eye strain. In sum, these effects hinder learning success. Future research is needed to clarify the underlying associations [17] to recommend designing VLE that do not overload and distract the learner, preventing physiological problems and fostering acceptance and learning performance. Guidelines and design principles aim to assist in developing a conducive and supportive VLE [6, 11] based on the learner's CL.

Cognitive Load in the Design of Learning Environments

The cognitive load theory is a largely validated principle [9] based on human cognitive architecture. CL refers to the mental capacity of a human. It describes that human information processing is restricted due to the limited capacity of the working memory. Since in the learning process, the most cognitive ability should be on the learning content, unnecessary information unrelated to learning or irritating instructions should be reduced to avoid any source that causes overload. The latest assumptions on cognitive load theory distinguish three dimensions of CL: *intrinsic cognitive load* (ICL), *germane cognitive load* (GCL) and *extraneous cognitive load* (ECL) [9].

The ICL is a more individual-related, stable dimension based on existing experience and knowledge inside the learner, as pre-existing expertise and knowledge influence how new information is processed. The GCL describes the CL as necessary for a successful learning process, defined as linking new content from working memory with each other and with knowledge from long-term memory. The ECL is mainly influenced by external stimuli like the instructional design [6, 9]. The ECL can be reduced by designing clear learning instructions and reducing unnecessary and redundant information. Based on that, design principles of learning environments aim to reduce CL since novel information must be learned efficiently without overstraining the learner [11, 18].

Many studies have confirmed the link between a decrease in the learner's ECL and better learning performance for analogue and digital learning [6, 9]. Options for using virtual reality for learning promote designing VLE that supports the learner to gain and apply knowledge in a new way. However, new virtual reality features have aspects beyond the existing features of online and digital learning, like immersion, movement and interactivity. A paradox arises from the desire to build realistic and highly immersive learning environments compared to well-established theories of CL [19].

This experiment investigates the relationship between immersion and interactivity of the learning material, CL and learning performance in two steps. Therefore, a VLE is designed and developed in an iterative, collaborative approach.

First, the effect of the degree of immersion of the VLE on CL is tested. For this, the degree of technical immersion is manipulated on two levels. In a between-subject design, two groups are compared, one learning in high immersion with virtual reality glasses (HMD) and one in low immersion with desktop virtual reality, hypothesising that:

(H1a): Learning in the highly immersive vle leads to a higher cognitive load than learning in the low immersive VLE.

Secondly, the effect of the interactivity of the learning material (on four levels) on the CL is tested. Learning material describes the presentation of the learning content. The level of interaction of the learning material is tested within-subject, meaning that one learner experiences all four levels from low interactive (audio and video) to high interactive (movement, 3D simulations), hypothesising that:

(H1b): Learning with a higher level of interactivity of the learning material leads to a higher cognitive load than learning with a lower level of interactivity.

The second step addresses the subquestion of an association between the degree of immersion, the CL and the learning performance.

(2a) Is there an association between the degree of immersion, the learner's cognitive load and the learning performance?

(2b) Is there an association between the four levels of interaction of the learning material, the learner's cognitive load and the learning performance?

Cognitive Load Measurement

The CL can be measured differently (objective and subjective) and on different dimensions (total CL and on the three subdimensions, ICL, ECL, GCL). Objective measures base their assessment on behavioural and physiological data, like eye tracking, EEG, and electrodermal or heart activity. Techniques like EEG, fMRI and eye-tracking can reflect CL pupil dilation, blink rate, fixation, and saccades [14–16].

In comparison, subjective measurements refer to information from questionnaires or self-reports. Commonly used questionaries are the Nasa-TLX [20] that differentiate the cognitive demand into different subscales (mental demand, physical demand, temporal demand, performance, effort and frustration) and questionnaires that assess the CL in its three dimensions ICL, ECL, GCL, e.g. [21–23].

For evaluating learning environments, the CL is often tested with self-report measurements at the end of the learning process [24–26]. Using validated questionnaires is a suitable option. However, measuring CL at the end or at certain time points during learning gives one or certain general indicators of CL. This approach does not explain human cognition's real-time response to a specific feature or learning material during the learning [27]. Research shows differences among measurements [28] but also for different time points [29]. Since it gets easier to measure eye-tracking with virtual reality or augmented reality glasses, it gets more common to use physiological assessments to measure CL.

This study uses eye-tracking to gain insights into the learner's CL response during learning. The following section gives an overview of the methodology to test the effects of degree immersion of a VLE and different levels of interactivity of learning materials on the learner's CL.

3 Material and Methods

Study Design

A between-subject design on the factor immersion (low and high immersion) with a within-subject design on the level of interaction of the learning material is chosen to address the research questions. The participants are randomly assigned to the immersion groups and perform a learning course, with four submodules presenting learning material in different interaction levels.

Each submodule differs in the learning material, like videos, podcasts, 3D animations and interactive 3D animations and their combinations (see Table 1). Depending on the learning material, the submodules differ in their interactivity level: A *shallow level* leaves the learner only a little opportunity for interaction by clicking only one button to activate videos and podcasts. At a *low level*, podcasts are linked to a content-relevant 3D model that supported the learning content. The learner can rotate the 3D model 360°. At a *high level* of interaction, movement is added to 3D animations,

Table 1 Learning material per submodule in the VLE

Submodule	Learning material	Learning content	Degree of interaction	Description
1	3D models and podcast	Definition and examples of rotary cutter glue parts	Low	360° rotatable 3D models
2	Video and podcast	How to produce rotary cutter glue parts	Very low	One pushable button to start the video
3	3D animations and podcast	The rotary cutter is an example for particular machine building	High	Stoppable and zoomable 3D animations
4	Interactive 3D animations and podcast	The rotary cutter in action	Very high	Changeable, stobable and zoomable 3D animations, exploded view

which are also combined with podcasts. Furthermore, the learner can stop the animation and zoom into the content-relevant 3D model. On a *very high level* of interaction, the learner can actively control the 3D animation and podcast content and display different production processing steps. They can also change the content-relevant 3D models into an exploded view with individual zoomable elements. Table 1 overviews the learning material used in each submodule. After each submodule, participants answer a knowledge test on the presented learning content.

Eye-tracking data will be collected during the whole learning for CL measurement with a primary interest in average pupil dilation and the variation of pupil dilation.

Learning Environment and Learning Material

The participants perform a learning course with knowledge of the tasks and activities of an adhesive technician, a profession from the adhesive industry. The course is designed as a VLE and is subdivided into four submodules. Each submodule presents unique learning content and uses different learning materials on four levels of interaction that are depicted in Table 1. For example, Fig. 1 shows the used 3D model of the rotary cutter with a high degree of interaction.

The VLE was created by university scientists and employees of a small engineering company collaboratively implemented in a research project funded by The Federal Ministry of Education and Research Germany (BMBF) and the European Social Fund. The main goal was to develop a VLE that contains a self-study workshop for machine operators in tape-converting enterprises (converters). The didactical concept is based on a competence-oriented learning journey following the development process of a single workpiece. Each submodule represents one production step of the workpiece: *From start to end—Rotary cutter glue parts (submodule 1)*,

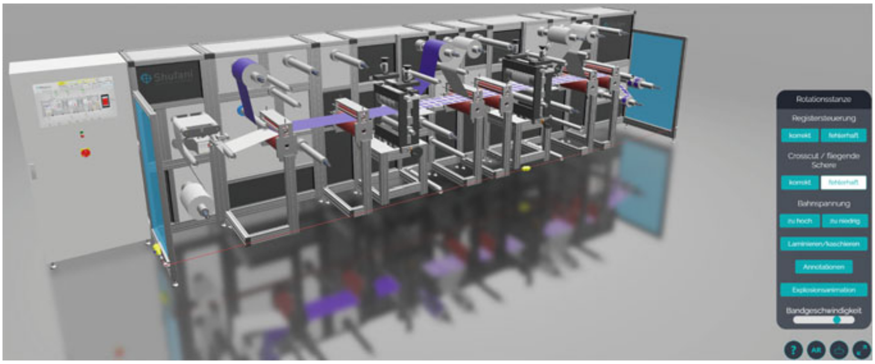


Fig. 1 3D model of the rotary cutter with buttons to start different configurations

Producing rotary cutter glue parts (submodule 2), *Special machine building—The rotary cutter (submodule 3)*, and *The rotary cutter in action (submodule 4)*. For every submodule, specific learning units have been defined and fruitful learning media selected and implemented following the guidelines of Mayer [6] for the design of VLE: Slides, videos, sounds, and/or animated 3D models. The design team also created relevant learning content based on the experience and knowledge provided by the experts of the engineering company. An additional onboarding area was created and enriched with a welcoming podcast and visual guidance to facilitate the learner's orientation before starting into the self-learn environment.

Procedure

Participants are students and employees of Potsdam University, located in Brandenburg, Germany. The participants are recruited with advertisements during lectures and flyers. Participation in the experiment is voluntary and not compensated. The data acquisition starts in February 2023.

Before starting the experiment, the participants are informed and asked to agree on the conditions of participation. They have the option to drop out at any time. Initially, a questionnaire assesses demographic variables and questions on prior expertise with virtual learning and gaming. Afterwards, the participants start with the experiment. The participants are randomly assigned to one experimental condition, half participating in the desktop condition (low immersion) and half in the virtual reality HMD condition (high immersion). In the desktop condition, the participants are equipped with the eye tracker, calibrated before starting the experiment. In the virtual reality HMD condition, the participants get the HMD glasses put on and calibrated.

Each participant has 30 min to move autonomously through the VLE (Fig. 2). In an additional onboarding area, the participants are introduced to the topic in general and to the structure of the VLE with a podcast and visual guidance (orientation map)

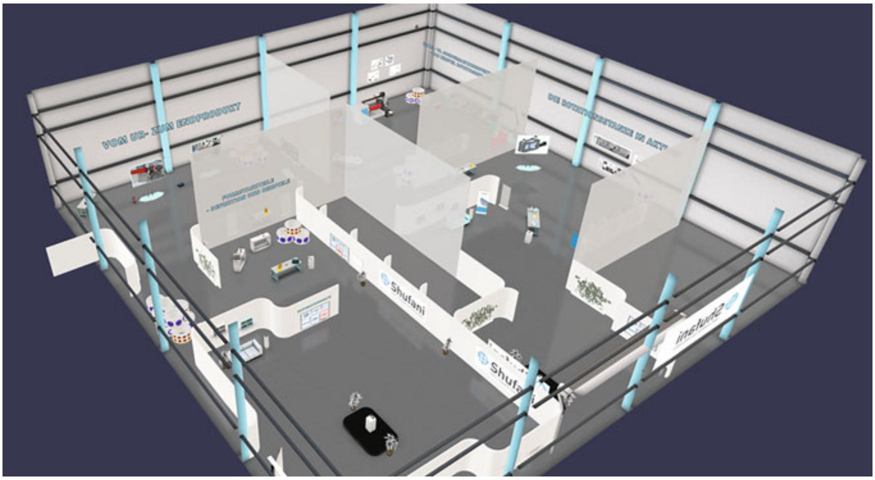


Fig. 2 The structure of the whole VLE and four learning submodules

to facilitate the learner's orientation before starting into the self-learn environment. After the onboarding area, the participants enter the first submodule. At the beginning of each submodule, the learning goals are presented. After each submodule, a knowledge test in the VLE tests the knowledge gained by the participants before they get automatically guided into the next submodule.

During the whole experiment, eye-tracking data is collected. After learning, a second questionnaire assesses the CL with a self-report questionnaire and the participants' perceived immersion and usability.

Measurements

Before and after the learning, two questionnaires are presented in German language and embedded online in *LimeSurvey*. After assessing demographic variables (age, gender, years of work experience, highest qualification), the participants are asked to briefly explain prior expertise with a virtual learning environment, virtual reality, or virtual gaming.

During the experiment, eye-tracking data is collected using the lightweight wearable system *ETVision* for measuring binocular points of gaze 180 times per second. The primary interest is on the parameters of average pupil dilation and variation of pupil dilation [14–16].

Knowledge is measured via a knowledge test after each submodule. Due to the topic's specificity, the participants are expected to have no prior experience with the learning content. For this reason, only a post-knowledge test is performed to avoid priming the participants for relevant content before the learning journey begins. The knowledge test was developed in collaboration with three individuals responsible for further educating individuals in the subject area of the learning content presented. The knowledge test consists of twenty multiple-choice questions covering all topics presented in the learning environment (e.g. *Which elements can be directly assigned to the cut-off machine? (a) Jumbo roll, (b) Cutting parts, (c) Distance box, (d)) Laminating station).* The maximum score on the knowledge test is 20 points, with 5 points for each submodule.

After the learning exposure, the learner's CL is also measured with a questionnaire using the scale of Andersen and Makransky [22]. With eight items, the *Cognitive Load Scale* assessed the subscales of ECL, ICL and GCL. As control variables, the data analysis will be enriched by measuring the learner's usability and perceived immersion via a questionnaire. The usability is measured with the *SUS Scale* [30], and the perceived immersion with Brooke [31].

Data Analysis

RStudio (Version 2022.12.0+353) will be used for statistical analysis. Data preparation and analysis of the eye-tracking recording will be performed with *ETV Vision.* Data from both eyes will be averaged to a single indicator for further analysis with a primary interest in average pupil dilation and variation of pupil dilation. Calculations of descriptive statistics and association will be performed for the following results: eye-tracking parameters (average of pupil dilation during saccade and pupil dilation during fixation and pupil dilation variation during saccade and pupil dilation variation during fixation following [16]), the results in the knowledge test, the CL via questionnaire after the learning phase, and usability and perceived immersion in total and for between-group and within-group comparison.

To answer the hypothesis, eye-tracking data will be compared statistically between the groups (H1a) and within one person (H1b).

4 Discussion

The cognitive load theory states that memory capacity is limited, and learning is enhanced by keeping enough resources for learning. Transferred to the design of learning environments, this postulates to design of clean and simple instructions. Recently, authors claimed that the assumptions on the relationship between CL and learning performance remained unanswered or need to be reconfirmed for virtual or augmented learning [9, 12]. Therefore, we performed an experimental study to clarify the relationship between CL and learning for 2° of immersive learning and to address research question one *Do a higher degree of immersion in the VLE and a higher interactivity of the learning material increase the learner's cognitive load?* Furthermore, research question two *Is there an association between immersion, interactivity of the learning material, and cognitive load with learning performance?*

To properly design VLE is highly relevant in terms of learning and education. In every part of daily life, the metaverse and opportunities for lifelong learning, equal education possibilities and borderless learning are on the rise [32]. New virtual worlds are being created before it is clear how they need to be designed so that employees, trainees, pupils and students can use them for learning. However, adverse effects such as cybersickness and mental fatigue make it clear that there is a need to catch up here. This experimental setting seeks empirical insights by focusing on the CL as one predictor for learning performance on the one hand and an indicator for cognitive overload on the other hand, using an eye-tracker to assess the cognitive response to immersive learning continuously.

Acknowledgements This research is part of the research project API-KMU (02L19A010 funded by the German Federal Ministry of Education and Research (BMBF). This research was also funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Project "German Internet Institute" (16DII137).

References

1. Alzahrani, N. M. (2020). Augmented reality: A systematic review of its benefits and challenges in e-learning contexts. *Applied Sciences, 10*(16), 5660. https://doi.org/10.3390/app10165660
2. Huang, W., Roscoe, R. D., Johnson-Glenberg, M. C., & Craig, S. D. (2021). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning, 37*(3), 745–758. https://doi.org/10.1111/jcal.12520
3. Klingenberg, S., Fischer, R., Zettler, I., & Makransky, G. (2022). Facilitating learning in immersive virtual reality: Segmentation, summarizing, both or none? *Journal of Computer Assisted Learning*, jcal.12741. https://doi.org/10.1111/jcal.12741
4. Lampropoulos, G., Keramopoulos, E., Diamantaras, K., & Evangelidis, G. (2022). Augmented reality and gamification in education: A systematic literature review of research, applications, and empirical studies. *Applied Sciences, 12*(13), 6809. https://doi.org/10.3390/app12136809
5. Liu, R., Xu, X., Yang, H., Li, Z., & Huang, G. (2022). Impacts of cues on learning and attention in immersive 360° video: An eye-tracking study. *Frontiers in Psychology, 12*, 792069. https://doi.org/10.3389/fpsyg.2021.792069
6. Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction, 29*, 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003
7. Breves, P., & Stein, J.-P. (2022). Cognitive load in immersive media settings: The role of spatial presence and cybersickness. *Virtual Reality*. https://doi.org/10.1007/s10055-022-00697-5
8. Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human-Computer Interaction, 38*(9), 801–824. https://doi.org/10.1080/10447318.2021.1976509
9. Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
10. van Merriënboer, J. J. G., Jelsma, O., & Paas, F. G. W. C. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development, 40*(2), 23–43. https://doi.org/10.1007/BF02297047

11. Bannert, M. (2002). Managing cognitive load—Recent trends in cognitive load theory. *Learning and Instruction, 12*(1), 139–146. https://doi.org/10.1016/S0959-4752(01)00021-4
12. Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, *34*(1), 171–196. https://doi.org/s
13. Hsieh, Y.-H., Lin, Y.-C., & Hou, H.-T. (2016). Exploring the role of flow experience, learning performance and potential behavior clusters in elementary students' game-based learning. *Interactive Learning Environments, 24*(1), 178–193. https://doi.org/10.1080/10494820.2013.834827
14. Rao, H. M., Smalt, C. J., Rodriguez, A., Wright, H. M., Mehta, D. D., Brattain, L. J., Edwards, H. M., Lammert, A., Heaton, K. J., & Quatieri, T. F. (2020). Predicting cognitive load and operational performance in a simulated marksmanship task. *Frontiers in Human Neuroscience, 14*, 222. https://doi.org/10.3389/fnhum.2020.00222
15. Shadiev, R., & Li, D. (2023). A review study on eye-tracking technology usage in immersive virtual reality learning environments. *Computers & Education, 196*, 104681. https://doi.org/10.1016/j.compedu.2022.104681
16. Shojaeizadeh, M., Djamasbi, S., Paffenroth, R. C., & Trapp, A. C. (2019). Detecting task demand via an eye tracking machine learning system. *Decision Support Systems, 116*, 91–101. https://doi.org/10.1016/j.dss.2018.10.012
17. Mayer, R. E., Makransky, G., & Parong, J. (2022). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human–Computer Interaction*, 1–10. https://doi.org/10.1080/10447318.2022.2108563
18. Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories.
19. *Educational Psychology Review*, *21*(1), 43–54. https://doi.org/10.1007/s10648-008-9096-1
20. Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007
21. Hart, S. G. (1986). *NASA task load index (TLX).*
22. Andersen, M. S., & Makransky, G. (2021). The validation and further development of the multidimensional cognitive load scale for physical and online lectures (MCLS-POL). *Frontiers in Psychology, 12*, 642084. https://doi.org/10.3389/fpsyg.2021.642084
23. Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997
24. Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1
25. Albus, P., & Seufert, T. (2022). Signaling in 360° desktop virtual reality influences learning outcome and cognitive load. *Frontiers in Education, 7*, 916105. https://doi.org/10.3389/feduc.2022.916105
26. Krüger, J. M., & Bodemer, D. (2022). Application and investigation of multimedia design principles in augmented reality learning environments. *Information, 13*(2), 74. https://doi.org/10.3390/info13020074
27. Lin, H.C.-S., Yu, S.-J., Sun, J.C.-Y., & Jong, M. S. Y. (2021). Engaging university students in a library guide through wearable spherical video-based virtual reality: Effects on situational interest and cognitive load. *Interactive Learning Environments, 29*(8), 1272–1287. https://doi.org/10.1080/10494820.2019.1624579
28. Bläsing, D., & Bornewasser, M. (2021). Influence of increasing task complexity and use of informational assistance systems on mental workload. *Brain Sciences, 11*(1), 102. https://doi.org/10.3390/brainsci11010102
29. Naismith, L. M., Cheung, J. J. H., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education, 49*(8), 805–814. https://doi.org/10.1111/medu.12732

30. Cheng, L.-P., Chang, L., Marwecki, S., & Baudisch, P. (2018). iTurk: Turning passive haptics into active haptics by making users reconfigure props in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–10). https://doi.org/10.1145/3173574.3173663
31. Brooke, J. (1995). SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, *189*.
32. Georgiou, Y., & Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human-Computer Studies, 98*, 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014
33. Ritterbusch, G. D., & Teichmann, M. (2023). Defining the Metaverse: A systematic literature review. In *IEEE Access* (Vol. PP). https://doi.org/10.1109/ACCESS.2023.3241809

Friend or Foe? Conversational Agents in the Digital Workplace and Their Effect on Users' Stress

R. Stefan Greulich, Magdalena Wiekenborg, Sascha Lichtenberg, Alfred Benedikt Brendel, and Milad Mirbabaie

Abstract Stress in the workplace and the resulting disorders present a significant challenge for the employee, employer, and national economy. According to the transactional model of stress, a possible reduction in stress can be achieved by reevaluating the situation. According to the social response theory, conversational agents should be very suitable for this task, as their increased persuasiveness leads to increased confidence of the user in their abilities. However, the perception of a social agent could also present an additional source of social evaluation stress, i.e., feeling observed and judged by another person. We propose a NeuroIS experiment utilizing EEG, ECG, and eye-tracking to resolve those opposing predictions.

Keywords Conversational agents · Work environment · Stress · Anthropomorphism

R. S. Greulich (✉) · S. Lichtenberg · A. B. Brendel
Chair of Business Information Systems, Esp. Intelligent Systems and Services, Technische Universität Dresden, Dresden, Germany
e-mail: stefan.greulich@tu-dresden.de

S. Lichtenberg
e-mail: sascha.lichtenberg@tu-dresden.de

A. B. Brendel
e-mail: alfred_benedikt.brendel@tu-dresden.de

M. Wiekenborg
Chair of Biopsychology, Technische Universität Dresden, Dresden, Germany
e-mail: magdalena.wekenborg@tu-dresden.de

M. Mirbabaie
Management Information Systems, Paderborn University, Paderborn, Germany
e-mail: milad.mirbabaie@upb.de

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_30

1 Introduction

The modern work environment is marked by increased mental stress, such as technostress [1], and the accompanying negative results for all stakeholders, including the employee, employer, and economy [2]. Mental health problems are cited as the second most prevalent cause of workplace absence, which is dominated by stress-induced disorders such as depression, anxiety, or burnout [3, 4]. Stress also increases the risk for other illnesses, such as cardiovascular, neurological, digestive tract, and even infectious diseases such as influenza, by weakening the immune system [5–8].

According to the transactional model of stress [9], a reduction in experienced stress can be achieved by consciously reevaluating the current situation. For instance, instead of focusing on the negatives, a problem can be reframed as an opportunity for personal growth [10]. In the end, to achieve this destressing, the subject must be encouraged that they have the necessary resources to deal with the situation or that it is not a threat.

A conversational agent (CA) that provides the impulse to reevaluate the situation is a possible solution for this challenge. According to the social response theory, humans experience increased perceived persuasiveness when interacting with a human-like CA [11]. This increased persuasiveness of a human-like CA would help users reevaluate the situation and gain confidence in their abilities. However, this perception of the CA as a social entity could also be a source of stress via the social evaluation threat [12], i.e., users feel observed and judged by the CA. We propose the following research question to resolve those opposing predictions:

How does the human-like design of a conversational agent influence the effect of destressing messages in a work-like environment?

To answer this question, we propose a NeuroIS experiment combining multiple physiological parameters to assess the user's reaction to a destressing CA in a work-like environment and its effects on experienced stress level and performance.

2 Theoretical Background

Transactional Stress Model by Lazarus

Stress is the body's physiological reaction to certain environmental stimuli, termed stressors [13]. The evolutionary function of the stress response is the activation of short term resources to deal with and survive threatening situations. As useful as this response is in the short-term, prolonged activation of the stress response carries a range of negative results, such as burnout, depression [14], anxiety [15], sleep disorders [16], increased risk of heart attack [17], and a reduced immune response [16].

In the transactional model of stress and coping [9], stress results from a multistage evaluation of external stimuli. The individuum interprets the stimuli as either positive, irrelevant, or dangerous. While irrelevant stimuli are discarded, dangerous stimuli are then evaluated to determine whether the individuum perceives their resources as adequate to deal with them. If that evaluation is negative, the user experiences stress.

Based on this theory, two main ways of reducing stress exist. Either the perception of threat is reduced or the user is encouraged that their resources are sufficient.

Conversational Agents and Stress

CAs are computer interfaces that interact with the user in a conversation-like manner utilizing written or spoken language [18]. CAs with human-like features, such as a name [19], an avatar [20], dynamic response delays [21], and the use of emoticons [22], elicit behavior in the user that is similar to human–human interaction. This has been described in the social response theory [23]. According to this theory, humans subconsciously perceive a CA as a social partner and act accordingly. Leading to the user experiencing higher enjoyment [24] and trust in the system [25]. Most importantly, it increases the perceived persuasiveness and leads to an increase in following behavioral recommendations made by the CA [11, 26, 27]. Hence, CAs exhibiting human-like design elements should provide stronger destressing by presenting messages in line with the transactional stress model.

However, increasing the perception of interacting with a social partner might lead to additional stress. As social beings, humans react strongly to being observed and evaluated by other humans [12]. This effect is termed social evaluation threat and has been shown to have an impact in various studies [28, 29]. If the CA is perceived as a social entity, it should also be seen as a social evaluation threat. In conclusion, combining social response theory and the transaction stress model leads to opposing predictions.

3 Hypotheses Development

As described in the previous section of this paper, the considered theories lead to opposing predictions regarding the effect of a human-like design on a CA presenting destressing messages (Fig. 1). We, therefore, postulate a unidirectional hypothesis: *(H1) An increase in human-like design leads to an increase or decrease in experienced stress by the user (social evaluation stress hypothesis or persuasive destressing hypothesis).*

An IS system, such as a CA, can be a source of technostress [1, 30–32]. We include a condition without a CA as a control condition and postulate that: *($H1_{con}$) The presence of a CA leads to an increase in experienced stress by the user*.

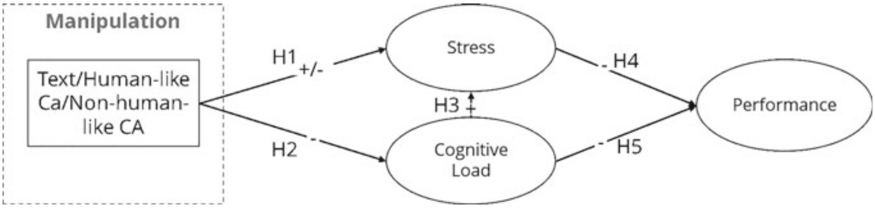


Fig. 1 Research Model. Stress was manipulated by providing different stressors

Cognitive load describes the differences between the user's mental capacities and task demands [33, 34]. Schmidhuber et al. compared the cognitive load when using a CA or a traditional software interface and found a reduction in cognitive load for the CA users [35]. Similarly, a virtual embodied assistant reduces cognitive load compared to an assistant using just a disembodied voice [36]. Therefore, we hypothesize: *(H2) An increase in human-like design leads to a decrease in the cognitive load of the user.*

According to the transactional stress model, stress is caused when the subject considers the task to exceed their mental resources [9]. In this context, an increase in cognitive load should lead to an increase in perceived stress: *(H3) Increased cognitive load leads to an increase in experienced stress by the user.*

We define the performance of the user as the accuracy with which the user is fulfilling the task. The impact of excessive stress on work quality is well documented [37–40] and should be reflected in our experiment: *(H4) The experienced stress decreases the performance of the user.*

Cognitive load has been shown to decrease the accuracy of complex tasks such as math problems [41–44]. We assume the same will hold true for our task and therefore postulate: *(H5) The experienced cognitive load decreases the performance.*

4 Method

Participants

We plan to collect data in an office-like setting from 90 to 100 right-handed participants recruited from students at a large German university (namely the TU Dresden).

Fig. 2 Workstation layout with the CA on the left and image labeling task on the right

Task and Procedure

Participants are instructed to label images of parked e-scooters, whether they are parked correctly, incorrectly, or if no scooter is depicted (Fig. 2). We structured the experiment into four five-minute blocks.

Treatments

The experiment has a mixed between- and within-subject design. With the stressors manipulated within-subject and the CA design manipulated between-subject. We will utilize a CA that communicates in written language, also termed a chatbot [45].

The stressors are manipulated within-subject with a block of 5-min labeling without the stressors, followed by two blocks with stressors and a last block without stressors. We selected the following stressors: time pressure via a clock counting the 5-min time backwards (similar to [46, 47]), street noise played over speakers (similar to [48]), and increased loading times of the images (similar to [49]). The CA will not initiate an interaction with the user during the image labeling but will provide a refresher of the task instructions upon request so as not to be a source of stress by itself.

We manipulate the chatbot design between subjects. One-third of the users will interact with a chatbot exhibiting human-like design characteristics, namely a name [19], self-reference [20], a dynamic typing delay [21], and utilizing emoticons in the conversation [22]. A third with a non-human-like designed chatbot, and one-third fulfill the task without interacting with a chatbot to provide a baseline for our experiment. Here, all necessary information is provided via a static text display (Table 1).

Table 1 Research design

Message	Human-like chatbot	Non-human like chatbot	Static text display
Greeting	Hello, I am Amanda your helpful work assistant	Hello, this is your helpful work assistant	<Missing>
Destressing message	Usually, people have no problem labeling enough images	Usually, people have no problem labeling enough images	Usually, people have no problem labeling enough images

Measures

For our NeuroIS experiment, we aim to complement questionnaires with physiological measurements [50, 51]. To assess the cognitive load of the participant, we will collect 16-channel EEG (OpenBCI) during the experiment. The cognitive load is measured by an increase in the alpha and theta ERD/ERS index over the whole cortex compared to a 5-min fixation baseline [52]. Physiological stress will be assessed using vagally-mediated heart rate variability (vmHRV, [53]). Without scientifically standardized state-of-the-art methodical practice for calculating vmHRV measures in NeuroIS [54], vmHRV will be operationalized via the mean root square of successive differences of RR intervals (RMSSD) calculated for each period from 2 min of continuous data. Additionally, we will collect eye-tracking data (GP3 HD from Gazepoint) to investigate how the user interacts with the presented display. Also, we will collect prior exposure to CAs with a modified technology exposure questionnaire [55] and demographic data to correct possible confounds in our experiment.

5 Discussion and Conclusion

In this research-in-progress paper, we presented an experiment investigating the effects of a destressing CA. We aim to resolve the opposing predictions arising from social response theory and the transaction stress model. Our study would provide valuable insight into the effects of CA in the digital workplace, stress prevention, and the overall impact of perceived humanness. Most importantly, if the social reaction to CAs only carries the positive effects of increased persuasiveness and work quality without drawback of being perceived as a social evaluator, destressing CAs could provide a valuable tool for workplace destressing. Theoretically, this result would show whether a CA with human-like design parameters is also perceived as a social evaluation threat.

Acknowledgements This work is supported by the Dr. Hans Riegel Foundation via the Connex Grant program (Grant Nr. 2207-025).

References

1. Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and validation. *Information Systems Research, 19*, 417–433. https://doi.org/10.1287/isre.1070.0165
2. DeVries, M. W., & Wilkerson, B. (2003). Stress, work and mental health: A global perspective. In: *Acta Neuropsychiatrica* (pp. 44–53). Cambridge University Press.
3. DAK-Gesundheit: DAK-Psychoreport 2019. (2019).
4. Knieps, F. (2021). BKK Gesundheitsreport 2021—BKK Dachverband.
5. Khansari, D. N., Murgo, A. J., & Faith, R. E. (1990). Effects of stress on the immune system. *Immunology Today, 11*, 170–175. https://doi.org/10.1016/0167-5699(90)90069-L
6. Lee, S. P., Sung, I. K., Kim, J. H., Lee, S. Y., Park, H. S., & Shim, C. S. (2015). The effect of emotional stress and depression on the prevalence of digestive diseases. *Journal of Neurogastroenterology and Motility, 21*, 273–282. https://doi.org/10.5056/jnm14116
7. Grant, I. (1985). The social environment and neurological disease. *Advances in Psychosomatic Medicine, 13*, 26–48. https://doi.org/10.1159/000411298
8. Steptoe, A., & Kivimäki, M. (2012). Stress and cardiovascular disease. *Nature Reviews. Cardiology, 9*, 360–370. https://doi.org/10.1038/nrcardio.2012.45
9. Lazarus, R. S., & Folkman, S. (1984) Stress, appraisal, and coping. In: *Health psychology: A handbook* (pp. 217–254).
10. Miller, T. A., & McCool, S. F. (2003). Coping with stress in outdoor recreational settings: An application of transactional stress theory. *Leisure Sciences, 25*, 257–275. https://doi.org/10.1080/01490400306562
11. Diederich, S., Lichtenberg, S., Brendel, A. B., Trang, S. (2020). Promoting sustainable mobility beliefs with persuasive and anthropomorphic design: Insights from an experiment with a conversational agent. In: *40th International Conference on Information Systems, ICIS 2019*.
12. Dickerson, S. S. (2008). Emotional and physiological responses to social-evaluative threat. *Social and Personality Psychology Compass, 2*, 1362–1378. https://doi.org/10.1111/j.1751-9004.2008.00095.x
13. Braceland, F. J. (1976). The stress of life. *Psychiatric Annals, 6*, 65–66. https://doi.org/10.3928/0048-5713-19761201-13
14. Iacovides, A., Fountoulakis, K. N., Kaprinis, S., & Kaprinis, G. (2003). The relationship between job stress, burnout and clinical depression. *Journal of Affective Disorders, 75*, 209–221. https://doi.org/10.1016/S0165-0327(02)00101-5
15. Charney, D. S., Deutch, A. Y., Krystal, J. H., Southwick, S. M., & Davis, M. (1993). Psychobiologic mechanisms of posttraumatic stress disorder. *Archives of General Psychiatry, 50*, 294–305. https://doi.org/10.1001/archpsyc.1993.01820160064008
16. Partinen, M. (1994). Sleep disorders and stress. *Journal of Psychosomatic Research, 38*, 89–91. https://doi.org/10.1016/0022-3999(94)90139-2
17. Bunker, S. J., Colquhoun, D. M., Esler, M. D., Hickie, I. B., Hunt, D., Jelinek, V. M., Oldenburg, B. F., Peach, H. G., Ruth, D., Tennant, C. C., & Tonkin, A. M. (2003). "Stress" and coronary heart disease: Psychosocial risk factors: National Heart Foundation of Australia position statement update. *Medical Journal of Australia, 178*, 272–276. https://doi.org/10.5694/j.1326-5377.2003.tb05193.x
18. McTear, M., Callejas, Z., & Griol, D. (2016). Conversational interfaces: Past and present. In: *The Conversational Interface* (pp. 51–72).
19. Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International Journal of Human-Computer Studies, 62*, 281–306. https://doi.org/10.1016/j.ijhcs.2004.11.008
20. Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2018). Designing anthropomorphic conversational agents: Development and empirical evaluation of a design framework. In: *International Conference on Information Systems 2018, ICIS 2018* (pp. 1–17). San Francisco.

21. Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2018). Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. In: *26th European Conference on Information Systems: Beyond Digitization—Facets of Socio-Technical Change, ECIS 2018* (p. 143975).
22. Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies, 132*, 138–161. https://doi.org/10.1016/J.IJHCS.2019.07.009
23. Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*, 81–103.
24. Lee, S. Y., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies, 103*, 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005
25. Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/J.CHB.2018.03.051
26. Brendel, A. B., Greve, M., Riquel, J., & Science, D. (2022). "Is it COVID or a cold ?" An investigation of the role of social presence, trust, and persuasiveness for users' intention to comply with COVID-19 Chatbots. In: *30th European Conference on Information Systems*, Timișoara, Romania (pp. 0–19).
27. Pietrantoni, N., Greulich, R. S., Brendel, A. B., & Hildebrandt, F. (2022). Follow me if you want to live—understanding the influence of human-like design on users' perception and intention to comply with COVID-19 education Chatbots. In: *Proceedings of the 43rd International Conference on Information Systems (ICIS)* (pp. 1–17).
28. Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The "Trier social stress test"—A tool for investigating psychobiological stress responses in a laboratory setting. In: *Neuropsychobiology* (pp. 76–81). Karger Publishers.
29. Kudielka, B. M., Hellhammer, H., & Kirschbaum, C. (2007). Ten years of research with the trier social stress test. In: *Social neuroscience* (pp. 56–83).
30. Riedl, R. (2012). On the biology of technostress. *ACM SIGMIS Database Base for Advances in Information Systems, 44*, 18–55. https://doi.org/10.1145/2436239.2436242.
31. La Torre, G., Esposito, A., Sciarra, I., & Chiappetta, M. (2019). Definition, symptoms and risk of techno-stress: A systematic review. *International Archives of Occupational and Environmental Health, 92*, 13–35. https://doi.org/10.1007/S00420-018-1352-1/TABLES/4
32. Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. In: *Information Systems Journal* (pp. 1010–1042). Blackwell Publishing Ltd.
33. Nachreiner, F. (1995). Standards for ergonomics principles relating to the design of work systems and to mental workload. *Applied Ergonomics, 26*, 259–263. https://doi.org/10.1016/0003-6870(95)00029-C
34. Sweller, J. (2011). Cognitive load theory. Academic Press.
35. Schmidhuber, J., Schlogl, S., & Ploder, C. (2021). Cognitive Load and Productivity Implications in Human-Chatbot Interaction. In: *Proceedings of the 2021 IEEE International Conference on Human-Machine Systems, ICHMS 2021*. Institute of Electrical and Electronics Engineers Inc.
36. de Melo, C. M., Kim, K., Norouzi, N., Bruder, G., & Welch, G. (2020). Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants. *Frontiers in Psychology, 11*, 3170. https://doi.org/10.3389/fpsyg.2020.554706
37. Ali, W. U., Raheem, A. R., Nawaz, A., & Imamuddin, K. (2014). Impact of stress on job performance: An empirical study of the employees of private sector Universities of Karachi. *Pakistan Journal of Management & Social Science, 3*, 14–17.
38. Bashir, U., & Ramay, M. I. (2010). Impact of stress on employees job performance a study on banking sector of Pakistan. *International Journal of Marketing Studies, 2*. https://doi.org/10.5539/ijms.v2n1p122.
39. Imtiaz, S., & Ahmad, S. (2009). Impact of stress on employee productivity, performance and turnover. An important managerial issue.

40. Pflanz, S. E., & Ogle, A. D. (2006). Job stress, depression, work performance, and perceptions of supervisors in military personnel. *Military Medicine, 171*, 861–865. https://doi.org/10.7205/MILMED.171.9.861
41. Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287–298. https://doi.org/10.1002/acp.1245
42. Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26*, 227–248. https://doi.org/10.1006/ceps.2000.1051
43. Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *American Journal of Psychology, 103*, 167. https://doi.org/10.2307/1423141
44. Deck, C., & Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review, 78*, 97–119. https://doi.org/10.1016/j.euroecorev.2015.05.004
45. Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. M. (2020). On the design of and interaction with conversational agents : An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems, 62*, Forthcoming. https://doi.org/10.1007/s12599-020-00639-y.
46. Buckert, M., Oechssler, J., & Schwieren, C. (2017). Imitation under stress. *Journal of Economic Behavior & Organization, 139*, 252–266. https://doi.org/10.1016/j.jebo.2017.04.014
47. Riedl, R., Kindermann, H., Auinger, A., & Javor, A. (2013). Computer breakdown as a stress factor during task completion under time pressure: Identifying gender differences based on skin conductance. *Advances in Human-Computer Interaction*. https://doi.org/10.1155/2013//20169.
48. Leather, P., Beale, D., & Sullivan, L. (2003). Noise, psychosocial stress and their interaction in the workplace. *Journal of Environmental Psychology, 23*, 213–222. https://doi.org/10.1016/S0272-4944(02)00082-8
49. Trimmel, M., Meixner-Pendleton, M., & Haring, S. (2003). Stress response caused by system response time when searching for information on the internet. In: *Human Factors* (pp. 615–621).
50. Tams, S., Thatcher, J., Hill, K., Grover, V., & De Guinea, A. O. (2014). NeuroIS—Alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems, 15*, 1. https://doi.org/10.17705/1jais.00374.
51. Fischer, T., & Riedl, R. (2017). Technostress research: A nurturing ground for measurement pluralism? Commun. *Association for Information Systems, 40*, 17. https://doi.org/10.17705/1CAIS.04017.
52. Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*, 425–438. https://doi.org/10.1007/s10648-010-9130-y
53. Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews, 36*, 747–756. https://doi.org/10.1016/J.NEUBIOREV.2011.11.009
54. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A Review of NeuroIS research with a focus on applied methods. In: Lect. Notes Inf. Syst. Organ. (Vol. 58, pp. 269–283). https://doi.org/10.1007/978-3-031-13064-9_28/TABLES/8.
55. Thompson, R. L., Higgins, C. A., & Howell, J. M. (1994). Influence of experience on personal computer utilization: Testing a conceptual model. *Journal of Management Information Systems, 11*, 167–187. https://doi.org/10.1080/07421222.1994.11518035

Heart Rate-Based Emotion Recognition and Adaptive Emotion Regulation Support with Wrist-Worn Wearables: A Systematic Literature Review

Elias Müller, Ivo Benke, and Alexander Maedche

Abstract Emotion regulation (ER) is a key skill since emotions play an essential role in personal development and in understanding social interaction. However, ER is unequally developed in humans and many are searching for ways to improve it. Neuro-adaptive systems bear a large potential for support in this endeavour with adaptive ER support based on biosignal. With the widespread use of wrist-worn wearables, new opportunities are emerging to capture biosignals, such as heart rate (HR), from the wearer in everyday life. This opens up the potential to use wrist-worn wearables to provide adaptive ER support. In this paper, we present a systematic literature review to provide an overview of the state-of-the-art in research on HR-based adaptive ER support with wrist-worn wearables. Specifically, we focus on the interplay between emotion recognition, wrist-worn wearables and adaptive ER support. Our findings show that HR-based wrist-worn wearable systems equally intervene via forms of feedback and external regulation by others and only four studies actually adapt the system. Further, many studies focus on response-based regulation or situation modification or selection strategies. To further research, we see research gaps in the ease of application of technical implementation, including biosignal processing and the use of ER support in systems.

Keywords Emotion recognition · Emotion regulation · NeuroIS · Biosignals · Adaptive systems

E. Müller (✉) · I. Benke · A. Maedche
Karlsruhe Institute of Technology (KIT), Kaiserstr. 89, 76133 Karlsruhe, Germany
e-mail: elias.mueller@kit.edu

I. Benke
e-mail: ivo.benke@kit.edu

A. Maedche
e-mail: alexander.maedche@kit.edu

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_31

1 Introduction

Emotion regulation (ER) is a key skill for humans since emotions play an essential role in personal development and in understanding social interaction [1, 2]. For example, emotions allow humans to assess specific situations, foster or hinder learning processes and memorize information [3]. Consequently, ER plays a central role in the development of transferable skills such as empathy, leadership, or self-empowerment [4]. ER refers to the things we do to influence [4] which emotions we have, when we have them, and how we experience and express them [5]. It is executed via the application of dedicated ER strategies [6, 7]. Consequently, the application of ER strategies is a key skill for developing transferable skills and for going through life successfully in general. However, ER is unequally developed in people and many are searching for ways to improve it and related skills as evidenced by the multitude of courses on personal development, e.g., in mindfulness, leadership, or empathy. In this paper, we take a look at the perspective on how to leverage sensor technologies embedded in wearables that support recognizing emotions via biosignals for ER support in the field. A promising electrical biosignal to understand emotions leverages electrocardiogram (ECG) to compute heart rate variability (HRV) [8]. Today, commercially available, affordable wearable devices provide the ability to capture cardiovascular activity. Specifically, smartwatches [9] as well as other wrist-worn wearables are becoming widely used in the population. According to Parks Associates, in 2021 22% of American households own a smartwatch which is most commonly used for activity, health and fitness tracking [10]. As wrist-worn wearables are equipped with sensors that provide increasingly accurate readings, the popularity of emotion-recognition smartwatches has also increased for research [11]. Although the sensor accuracy does not match other wearables [12], like a chest strap, wrist-worn wearables are being used in various studies and have shown to provide valuable data for recognizing emotions. By being non-invasive, widespread and suitable for everyday use, they offer interesting possibilities for use in neuroadaptive systems, which is one of the research areas of the NeuroIS community and is expected to grow in the future [13–15]. In contrast to existing publications delivered by the NeuroIS community that focused on specific sensors and biosignals [12, 16–19], the focus of our review is the interplay between (1) emotional states derived from biosignals and (2) the NeuroIS system using wrist-worn wearables.

To obtain an overview of the possible applications of wrist-worn wearables and their possible integration into emotion-adaptive systems, we conducted a systematic literature review (SLR) to provide a consolidated view of research findings on ER support using emotion-aware smartwatches answering the following research question: *What is the state-of-the-art in research on emotion-adaptive regulation support using wrist-worn wearables with heart rate data?*

We analyzed the publications identified in our SLR through a morphological box based on a conceptual model of adaptive ER support systems which consists of the steps emotion recognition, the adaptation logic, the intervention type of the adaptive ER support system, and the ER strategy family in place (see Sect. 3). Our results show that wrist-worn wearable systems based on biosignals from the heart rate (HR) intervene to support ER via feedback and notification, through external regulation by others. Only four studies actually adapt the system. Further, many studies focus on response-based ER or situation modification or selection strategies.

2 Method

In our paper, we followed the SLR search process proposed by Webster and Watson [20] to categorize the various data records. First, we defined the scope of the search in form of the search string. It consists of three concepts: wrist-worn wearables, the ability to recognize and regulate emotions, and the cardiac biosignals to recognize emotions. The detailed search string was: *("smart device" OR "smartwatch" OR "smart watch" OR wrist-band OR "wrist band" OR "wrist bands" OR "wrist worn" OR "wrist-worn") AND ("emotion regulation" OR (emotio* AND regulat*) OR emotion-aware OR "emotional context" OR "affect regulation" OR affect-aware OR "affective context") AND (heart-rate OR "heart rate" OR ecg).* Second, we selected the databases ACM digital library, IEEE Xplore, AIS eLibrary, Springer Link, and Web of Science Core Collection for our search. In the databases that allowed an abstract search, this was applied to the first and second "AND" condition of the search string. Third, we included only publications published later than 2014 since corresponding wearables smartwatches are a relatively innovative technology, e.g., the Apple Watch was introduced in 2014 [21]. Another inclusion criterion was that we only looked at peer-reviewed journals and conference papers. With our search strategy, we obtained 564 results. Through manual reviewing them on the SLR objective, we received 23 results in total. In a forward-backwards, we found additional 9 publications, whereby one publication targeted the same prototype, were found. In a final step, the 31 results were analyzed according to the studies conducted and the Conceptual Model explained in Sect. 3.

3 Conceptual Model

We conceptualize emotion recognition and adaptive ER support based on wrist-worn wearables with a cycle consisting of four steps (see Fig. 1).

Our conceptual model distinguishes between a human and technical system perspective. From a human perspective the central trigger for adaptation is the appearance of emotions. Emotions in general can either be assessed manually (e.g., through surveys) or automatically. We follow an automated approach to discover emotions. In step (1) **emotion recognition**, sensors (in our case mostly embedded wrist-worn wearables) are leveraged, raw data processing is performed and classification algorithms are applied. Subsequently, step (2) covers the **adaptation logic**. This step builds on the detected emotion and describes the procedure how an intervention will

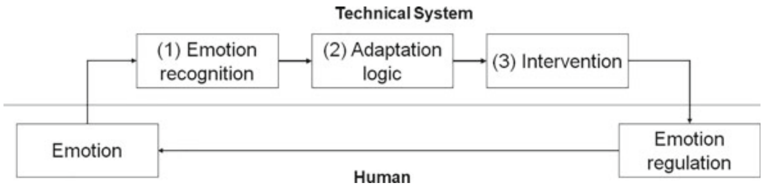


Fig. 1 Conceptual model

be triggered based on the recognized human's emotion. Finally, step (3) **intervention** describes how the system responds to the recognized emotion and the defined adaptation logic in order to influence human behavior. These interventions can be, for example proposing exercises to be performed by the human, adjustments to an application or feedback.

Finally, again from a human perspective, **emotion regulation** describes the strategy for regulating the human's emotion during or after an intervention. Existing literature typically suggests grouping of strategies in 5 families: Situation Selection, Situation Modification, Attentional Deployment, Cognitive Change and Response Modulation [22],

4 Results

We present the results of our SLR in a morphological box (Fig. 2), the underlying data is captured in Table 1 in the appendix. The dimensions of the morphological box describe the study characteristic, the emotion, the steps from the technical system, and the emotion regulation.

The dimension **study characteristics** provides information on the type of investigation conducted in the studies, either for data collection (in the field and in the laboratory, 7 in total) or for the assessment of artefacts. 15 studies were conducted in a laboratory setup and 10 studies in a field setup. In 6 studies, no or other types of studies were conducted.

With regards to the **emotional model**, 13 studies followed a discrete emotional model. Five studies followed a continuous dimensional emotion model (i.e., valence-arousal), in nine studies an emotion-centric construct was classified, for example, in [23] the user's laughter was recognized. In four studies no model was reported.

The dimension of the **discrete emotion** is differentiated from the upper dimension in more detail. Anger and disgust were each chosen once as a discrete emotion, anxiety seven times, joy and surprise twice and sadness three times. Although stress is not an emotion, we added it to this dimension because the intervention for emotions is also partially applicable to stress; it was used in seven studies. Among others (5), we included related states to the discrete emotions, such as laughter [23] or engagement [24].

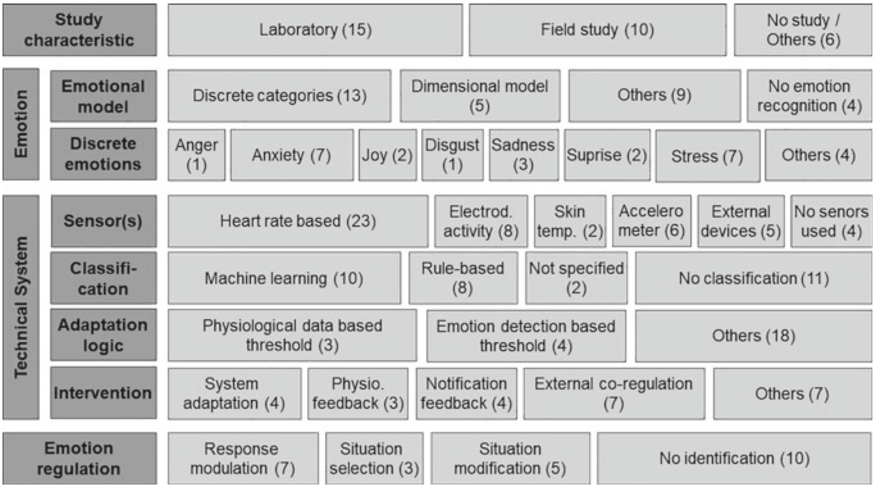


Fig. 2 Morphological box for the SLR analysis

The dimension of **sensors** covers the various physiological sensors that have been used by wearables, mostly wrist-worn wearables, used for emotion recognition and during adaption. 23 prototypes have measured the HR. Electrodermal activity sensors were used in total of eight times. Six times the accelerator and two times the skin temperature sensor was used. In five studies, additional measuring devices were used, which recorded data that could not been measured with a wrist-worn wearable like for example, Dong et al. [25] also used a camera to capture facial expressions. Four times there was no emotion recognition, such as in CoolCraig [26]. This system focuses on behaviour and emotions in children with ADHD in a co-regulation setting.

For the **classification** of the emotion using raw sensor data, machine learning was used in ten studies, rule-based classification in eight studies, no indication was given in two studies and no emotion classification took place 11 times. This included also prototypes using emotion-related biosignals, for example HeartChat [27], which developed an HR augmented chat for emotion regulation in chats. Another example is EmotionCheck [28], which in their laboratory study evaluated the effectiveness of the intervention and the anxiety was induced by the study.

For the following analysis, we excluded all studies (6) that did not conduct the support of emotion regulation but rather focused only on recognition. We continue with those studies that represent the whole conceptual model (25).

The **adaptation logic** was based on a threshold of sensor values or features in three studies, in four studies the logic was based on the detected discrete emotion, and the remaining 18 are summarised under others. These include studies that induced the emotion or that co-regulation took place.

Different **intervention types** were leveraged: In four studies, an automatic adaptation of the system takes place, three the user received haptic rhythmic feedback on their wrist. Four times an exercise was suggested to the user by a notification. In

Table 1 Results of the SLR

Dimension	Characteristic	References	#
Study characteristic	Laboratory study	[23, 28–30, 32–42]	15
	Field study	[24, 27, 31, 43–49]	10
	No study	[25, 26, 50–53]	6
Sensors	Heart rate based	[23, 25, 27–37, 43–49, 51–53]	23
	Electrodermal activity	[23, 24, 29, 36, 37, 40, 42, 42]	8
	Skin temperature	[37, 52]	2
	Accelerometer	[23, 29, 38, 39, 45, 46]	6
	External device	[25, 35, 37, 46, 47]	5
	No Recognition	[26, 30, 41, 50]	4
Classification	Machine learning	[23, 25, 29, 37, 38, 40, 45, 47, 48, 52]	10
	Rule-based	[24, 32, 33, 35, 43, 44, 51, 53]	8
	Not specified	[31, 34]	2
	No classification	[26–28, 30, 36, 39, 41, 42, 46, 49, 50]	11
Emotional model	Discrete categories	[28, 31–34, 38, 42–44, 48, 51–53]	13
	Dimensional model	[25, 29, 35, 40, 47]	5
	Others	[23, 24, 27, 36, 37, 39, 45, 46, 49]	9
	No emotion recognition	[26, 30, 41, 50]	4
Discrete emotions	Anger	[34]	1
	Anxiety	[28, 31–34, 42, 52]	7
	Joy	[38, 51]	2
	Disgust	[34]	1
	Sadness	[34, 38, 51]	3
	Surprise	[34, 40]	2
	Stress	[30–32, 43, 44, 48, 51]	7
	Others	[24, 37, 44, 45]	4
Adaptation logic	Physiological data based threshold	[32, 33, 44]	3

(continued)

Table 1 (continued)

Dimension	Characteristic	References	#
	Emotion detection based threshold	[25, 34, 41, 51]	4
	Others	[24, 26–28, 30, 31, 35, 36, 39, 40, 42, 43, 46, 48–50, 52, 53]	18
Intervention	System adaptation	[25, 32, 44, 51]	4
	Physiological feedback	[28, 33, 42]	3
	Notification feedback	[30, 31, 35, 46]	4
	External co-regulation	[24, 27, 36, 39, 40, 49, 53]	7
	Others	[26, 34, 41, 43, 48, 50, 52]	7
Emotion regulation	Response modulation	[28, 31–33, 35, 41, 42]	7
	Situation selection	[25, 43, 51]	3
	Situation modification	[24, 30, 36, 39, 44]	5
	No identification	[26, 27, 34, 40, 46, 48–50, 52, 53]	10

seven studies, adaptation took place through external co-regulation and seven were categorised under other.

Finally, the distribution among the **emotion regulation** strategy families were the following: The response modulation family was used in seven studies, the situation selection in three studies and the situation change strategy family in five studies. In ten studies we could not make an allocation.

5 Discussion

In this study we conducted a SLR to provide an overview of research on HR-based adaptive ER support with wrist-worn wearables. Based on the results of our SLR we identified two main streams in previous research.

First, the results show that 80% of the studies conduct and investigate ER support, while the other studies focused on emotion recognition using wrist-worn wearables. These studies represent the first stream. In this stream emotion recognition is mostly conducted in a multimodal approach with multiple sensors, post-processing, and classification with machine learning. Many of these studies applied a two-dimensional valence-arousal model to understand emotions. EmotionSence [29] is an example from this stream. In a lab study, they took data from different users in different activity states, such as sitting and walking, and built different models to detect emotions in a 2D model.

The second stream includes all studies that supported ER and included either an adaptation or an intervention. Mostly the focus was on evaluating the regulation intervention or adaptation concept and not emotion recognition, which was partly induced by the study and not automatically performed. Many of the prototypes in these studies were based on the discrete or categorical emotion models, of these a lot focused on negative emotions like anxiety, anger, or stress. The adaptation logic in these studies was quite different. About one fifth of the studies gave general feedback in the form of a message to users to recommend an exercise, like FOQUS [30], which includes a feature that helps to be focused by applying the Pomodoro time management technique [31].

Four studies adapted the interaction with or the presentation of the system by adjusting the speed of interaction or adaption of the system interface. In seven studies a special form of intervention was chosen in which the emotional state or physiological data was communicated to a co-regulator who then decided whether to change the mode of communication or recommended an exercise. In three studies, the user receives artificial, heartbeat-like rhythmic feedback on the wrist to subtly calm and regulate their HR. Finally, when looking at the ER strategies applied in the studies, three groups stand out. Seven studies focused on a response-based approach by elicitation of response modulation in the participants. For example, Di Lascio et al. [23] have developed a system called *FishBuddy* that detects stress and anxiety in learners during an exercise and interrupts the learning process to regulate the emotion by using a simulated fish whose swimming rhythm reduces the HR. Second, five studies regulated emotions by modifying the situation, and three started even earlier in the ER model by supporting the application of situations selection strategies. Other emotion strategies were not applied in the sample. A reason for this might be since the ER strategy attention deployment and cognitive change require a deeper understanding of the situation and the context in order to support ER through a neuro-adaptive system.

Limitations and Future Research

Although we followed a rigorous search approach our work comes with limitations. We had to make decisions with regards to the databases searched, the exclusion and inclusion criteria, and the analysis levels. We will address these aspects in future cycles. Based on our results and limitations, we have discovered potential for future research. Despite the prevalence of wrist-worn wearables in the population, we see a research gap in the area of adaptive ER support systems with an integration of all steps of our conceptual model such as adaptation logic, intervention, and regulation. Second, only three of the five ER strategy families are investigated in the studies.

To address these gaps, we see the potential for further research to bring together the technical implementation and processing of biosignals and ER support strategies and their use cases. This requires (1) methodological contributions that conceptually describe the interplay of emotion recognition, logic and intervention. Furthermore, we believe that a toolkit supporting the different steps of emotion recognition and adaptation through standardized software modules could be beneficial. This would make it possible to prototype ER support applications based on wrist-worn wearables for researchers from non-technical disciplines.

Appendix

See Table 1.

References

1. Ensor, R., Spencer, D., & Hughes, C. (2011). 'You Feel Sad?' Emotion understanding mediates effects of verbal ability and mother-child mutuality on prosocial behaviors: Findings from 2 years to 4 years. *Social Development, 20*, 93–110. https://doi.org/10.1111/j.1467-9507.2009.00572.x
2. Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*, 149–180. https://doi.org/10.1007/s10919-009-0070-5
3. Tyng, C. M., Amin, H. U., Saad, M. N. M., Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology 8*.
4. Gross, J. J., & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*, 8–16. https://doi.org/10.1177/1754073910380974
5. Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299. https://doi.org/10.1037/1089-2680.2.3.271
6. Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*, 1–26. https://doi.org/10.1080/1047840X.2014.940781
7. Gross, J. J. (2014). Emotion regulation: Conceptual and empirical foundations. In: *Handbook of emotion regulation* (2nd ed., pp. 3–20). The Guilford Press.
8. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*, 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010
9. Wearable Technology Market Share & Trends Report. (2030). Retrieved February 23, 2023, from https://www.grandviewresearch.com/industry-analysis/wearable-technology-market
10. Parks Associates—Smart Product Market Assessments. Retrieved March 10, 2023, from http://www.parksassociates.com/services/smart-product-assessments
11. Schmidt, P., Reiss, A., Dürichen, R., & Laerhoven, K. V. (2019). Wearable-based affect recognition—A review. *Sensors, 19*, 4079. https://doi.org/10.3390/s19194079
12. Stangl, F. J., & Riedl, R. Measurement of heart rate and heart rate variability with wearable devices: A systematic review.
13. Fischer, T., Davis, F. D., & Riedl, R. (2019). NeuroIS: A survey on the status of the field. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information systems and neuroscience* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-030-01087-4_1

14. vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a NeuroIS research agenda with four areas of societal contributions. *European Journal of Information Systems, 29*, 9–24. https://doi.org/10.1080/0960085X.2019.1708218
15. Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: Progress, challenges, and future directions. *SIGMIS Database, 51*, 13–54. https://doi.org/10.1145/3410977.3410980
16. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability: A review of NeuroIS research with a focus on applied methods. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M., Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 269–283). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_28
17. Stangl, F. J., & Riedl, R. (2022). Measurement of heart rate and heart rate variability in NeuroIS research: Review of empirical results. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 285–299). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_29
18. Seitz, J., & Maedche, A. (2022) Biosignal-based recognition of cognitive load: A systematic review of public datasets and classifiers. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 35–52). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_4
19. Dorner, V., & Uribe Ortiz, C. E. (2022) New measurement analysis for emotion detection using ECG data. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. B. Randolph, & G. R. Müller-Putz (Eds.), *Information systems and neuroscience* (pp. 219–227). Springer International Publishing. https://doi.org/10.1007/978-3-031-13064-9_23
20. Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*, xiii–xxiii.
21. Apple Unveils Apple Watch—Apple's Most Personal Device Ever. Retrieved March 4, 2023, from https://www.apple.com/newsroom/2014/09/09Apple-Unveils-Apple-Watch-Apples-Most-Personal-Device-Ever/
22. Sheppes, G., & Gross, J. J. (2012) Emotion regulation effectiveness: What works when. In I. Weiner (Ed.), *Handbook of psychology* (2nd ed., p. hop205018). Wiley. https://doi.org/10.1002/9781118133880.hop205018
23. Di Lascio, E., Gashi, S., & Santini, S. (2019). Laughter recognition using non-invasive wearable devices. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 262–271). Association for Computing Machinery. https://doi.org/10.1145/3329189.3329216
24. DiSalvo, B., Bandaru, D., Wang, Q., Li, H., & Plötz, T. (2022). Reading the room: Automated, momentary assessment of student engagement in the classroom: Are we there yet? In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 6, pp. 1–26). https://doi.org/10.1145/3550328
25. Dong, Y., Hu, H., Wen, Y., Yu, H., & Miao, C. (2018) Personalized emotion-Aware video streaming for the elderly. In *Social Computing and Social Media. Technologies and Analytics: 10th International Conference, SCSM 2018*, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II (pp. 372–382). Springer. https://doi.org/10.1007/978-3-319-91485-5_28
26. Doan, M., Cibrian, F. L., Jang, A., Khare, N., Chang, S., Li, A., Schuck, S., Lakes, K. D., & Hayes, G. R. (2020). CoolCraig: A smart watch/phone application supporting co-regulation of children with ADHD. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3334480.3382991
27. Hassib, M., Buschek, D., Wozniak, P. W., & Alt, F. (2017). HeartChat: Heart rate augmented mobile chat to support empathy and awareness. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2239–2251). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025758

28. Costa, J., Adams, A. T., Jung, M. F., Guimbretière, F., & Choudhury, T. (2017). EmotionCheck: A wearable device to regulate anxiety through false heart rate feedback. *GetMobile: Mobile Computing and Communications, 21*, 22–25. https://doi.org/10.1145/3131214.3131222
29. Wang, Z., Yu, Z., Zhao, B., Guo, B., Chen, C., & Yu, Z. (2020) EmotionSense: An adaptive emotion recognition system based on wearable smart devices. *ACM Transactions on Computing for Healthcare, 1*, 20:1–20:17. https://doi.org/10.1145/3384394
30. Dibia, V. (2016). FOQUS: A smartwatch application for individuals with ADHD and mental health challenges. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 311–312). Association for Computing Machinery. https://doi.org/10.1145/2982142.2982207
31. Morris, J. T., & Wallace, T. (2018). BreatheWell wear: Development and testing of a stress management app for TBI and PTSD on android wear smartwatches. In K. Miesenberger, & G. Kouroupetroglou (Eds.), *Computers helping people with special needs* (pp. 472–478). Springer International Publishing. https://doi.org/10.1007/978-3-319-94274-2_69
32. Chen, J., Zhu, B., Balter, O., Xu, J., Zou, W., Hedman, A., Chen, R., & Sang, M. (2017). FishBuddy: Promoting student engagement in self-paced learning through wearable sensing. In *2017 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 1–9). https://doi.org/10.1109/SMARTCOMP.2017.7947008
33. Costa, J., Guimbretière, F., Jung, M. F., & Choudhury, T. (2019). BoostMeUp: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 3, pp. 1–23). https://doi.org/10.1145/3328911
34. Vermeulen, J., MacDonald, L., Schöning, J., Beale, R., & Carpendale, S. (2016). Heartefacts: Augmenting mobile video sharing using wrist-worn heart rate sensors. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 712–723). ACM. https://doi.org/10.1145/2901790.2901887
35. Liu, N., Zhang, Y., Mark, G., Li, Z., & Rau, P.-L.P.: Mindfulness meditation: Investigating immediate effects in an information multitasking environment. In P.-L.P. Rau (Ed.), *Cross-cultural design. methods, tools and user experience* (pp. 531–542). Springer International Publishing. https://doi.org/10.1007/978-3-030-22577-3_38
36. Lee, M., Kim, K., Rho, H., & Kim, S. J. (2014). Empa talk: A physiological data incorporated human-computer interactions. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (pp. 1897–1902). Association for Computing Machinery. https://doi.org/10.1145/2559206.2581370
37. Salah, J., Abdelrahman, Y., Abdrabou, Y., Kassem, K., & Abdennadher, S. (2018). Exploring the usage of commercial bio-sensors for multitasking detection. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (pp. 265–277). Association for Computing Machinery. https://doi.org/10.1145/3282894.3282900
38. Quiroz, J. C., Yong, M. H., & Geangu, E. (2017). Emotion-recognition using smart watch accelerometer data: Preliminary findings. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 805–812). ACM. https://doi.org/10.1145/3123024.3125614
39. Joi, Y. R., Jeong, B. T., Kim, J. H., Park, K. H., Lee, T., & Cho, J. D. (2015). WearLove: Affective communication via wearable device with gamification. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 559–564). Association for Computing Machinery. https://doi.org/10.1145/2793107.2810337
40. Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 2, pp. 103:1–103:21). https://doi.org/10.1145/3264913
41. Williams, M. A., Roseway, A., O'Dowd, C., Czerwinski, M., & Morris, M. R. (2015) SWARM: An actuated wearable for mediating affect. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 293–300). Association for Computing Machinery. https://doi.org/10.1145/2677199.2680565

42. Azevedo, R. T., Bennett, N., Bilicki, A., Hooper, J., Markopoulou, F., & Tsakiris, M. (2017). The calming effect of a new wearable device during the anticipation of public speech. *Scientific Reports, 7*, 2285. https://doi.org/10.1038/s41598-017-02274-2
43. Hwang, C., & Pushp, S. (2018). A mobile system for investigating the user's stress causes in daily life. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (pp. 66–69). ACM. https://doi.org/10.1145/3267305.3267668
44. Szwoch, M., & Szwoch, W. (2019). Using different information channels for affect-aware video games—A case study. In M. Choraś, & R. S. Choraś (Eds.), *Image processing and communications challenges* (Vol. 10, pp. 104–113). Springer International Publishing. https://doi.org/10.1007/978-3-030-03658-4_13
45. Hafiz, P., Miskowiak, K. W., Maxhuni, A., Kessing, L. V., & Bardram, J. E.: Wearable computing technology for assessment of cognitive functioning of bipolar patients and healthy controls. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 4, pp. 129:1–129:22). https://doi.org/10.1145/3432219
46. Cambo, S. A., Avrahami, D., & Lee, M. L. (2017). BreakSense: Combining physiological and location sensing to promote mobility during work-breaks. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3595–3607). Association for Computing Machinery. https://doi.org/10.1145/3025453.3026021
47. Exler, A., Schankin, A., Klebsattel, C., & Beigl, M. (2016). A wearable system for mood assessment considering smartphone features and data from mobile ECGs. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 1153–1161). Association for Computing Machinery. https://doi.org/10.1145/2968219.2968302
48. Boateng, G., & Kotz, D. (2016). StressAware: An app for real-time stress monitoring on the amulet wearable platform. In: *2016 IEEE MIT Undergraduate Research Technology Conference (URTC)* (pp. 1–4). https://doi.org/10.1109/URTC.2016.8284068
49. Liu, F., Esparza, M., Pavlovskaia, M., Kaufman, G., Dabbish, L., & Monroy-Hernández, A. (2019). Animo: Sharing biosignals on a smartwatch for lightweight social connection. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Vol. 3, pp. 18:1–18:19). https://doi.org/10.1145/3314405
50. Zheng, H., & Genaro Motti, V. (2018). Assisting students with intellectual and developmental disabilities in inclusive education with smartwatches. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3173574.3173924
51. Choi, J., Lee, J., & Kim, H. (2017). Extending app pre-launch service with emotion context: Poster abstract. In *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation* (pp. 329–330). ACM. https://doi.org/10.1145/3054977.3057306
52. Miranda Calero, J. A., Rituerto-González, E., Luis-Mingueza, C., Canabal, M. F., Bárcenas, A. R., Lanza-Gutiérrez, J. M., Peláez-Moreno, C., & López-Ongil, C. (2022). Bindi: Affective Internet of Things to combat gender-based violence. *IEEE Internet of Things Journal, 9*, 21174–21193. https://doi.org/10.1109/JIOT.2022.3177256
53. Qin, C. Y., Choi, J.-H., Constantinides, M., Aiello, L. M., & Quercia, D. (2021). Having a heart time? A wearable-based biofeedback system. In: *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3406324.3410539

Design of an In-Class Virtual Calm Down Space for Neurodiverse Students

Jackson Ritchey, Yasmean Dehaghani, and Tamara Roth

Abstract Infia Studios has developed virtual reality (VR) "calm down spaces" for use in school classrooms. While aimed at helping students on the Autism Spectrum, the calming effects are available for all students. Developed as a SmartPhone app called "Serenity," the concept is to provide a better calming experience (vs a physical calm down space) for a distressed student while helping minimize disruption to learning. The data presented here is from a "proof of concept" demonstration study conducted at a local elementary school with 15 participants aged 9–11 years old. Each student experienced the Serenity VR app using a VR headset for three minutes, then was asked a brief series of questions. Without fail, all students reported a positive experience and a calming effect from the digital environments.

Keywords Neurodiversity · Virtual Reality · Calm-down Space · Classroom Technology · Digital Environments

J. Ritchey · Y. Dehaghani · T. Roth (✉)
Sam M. Walton College of Business, University of Arkansas, Fayetteville, USA
e-mail: tamara.roth@uni.lu; TRoth@walton.uark.edu

J. Ritchey
e-mail: JRitchey@walton.uark.edu

Y. Dehaghani
e-mail: ydehagha@uark.edu

T. Roth
FINATRAX, Interdisciplinary Center Security, Reliability and Trust (SnT), University of Luxembourg, Kirchberg, Luxembourg

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68,
https://doi.org/10.1007/978-3-031-58396-4_32

1 Introduction and Related Work

Through no fault of their own, over a million U.S. school-aged children struggle with dysregulated emotional events, which often manifest as outbursts that disrupt classroom learning [1]. The problem exists in classrooms of all levels, from kindergarten through graduate school, and affects a significant number of students across the country [2]. Dysregulated emotional outbursts are difficult for both the child experiencing it, the teacher in charge, as well as every other student in the classroom. Such events often bring learning to a halt while the teacher attends to the needs of the student [3].

Among students on the autism spectrum, anxiety-related emotional events are most prevalent [4]. Between 42 and 55 percent of students show behaviors characteristic for at least one anxiety disorder. The disorders range from (social) phobias to generalized anxiety disorder, obsessive–compulsive and separation anxiety disorder. They affect the wellbeing of students on the ASD spectrum across all levels of classroom learning, including cognitive functions [5]. In the classroom, anxiety typically results from sensory overload, the disruption of routines, or social interaction [4]. Teachers often respond to anxiety-related emotional events by either sanctioning or overprotecting students and avoiding triggers [6].

One common and contemporary solution attempted in countless schools is the "calm down space". A calm down space is a designated area in a home or classroom with the sole intent of being a quiet, safe space for a child to go to when they feel their emotions are running too high and they need to regain their emotional and physical control [7–9]. In schools with few resources, calm down spaces are the best-case scenario for students in the classroom. Anecdotally, most teachers report that calm down spaces are at best an inadequate half measure. Many teachers do not know much about the psychological requirements of such calm down spaces [4] and provide nothing more than a few scattered pillows for students. Many students, thus, perceive calm down spaces as a punishment not a solution to their problem.

Recent advances in technology innovation position virtual reality (VR) as an alternative to tackle anxiety-related emotional events. Most research focuses on the use of VR as a safe space for persons with anxiety disorders and phobias, where they can be confronted with triggers and learn how process their emotions in stressful situations [10]. Other research explores VR as a possibility to provide relaxation on negative and positive affective states. The immersive experience from auditory and visual stimuli may counteract negative emotional events and provide immediate relief [11, 12]. Building on this research, we see many opportunities for a virtual calm down spaces in the classroom. Thus, we ask the following the question:

RQ: *How can the use of information systems help create calm down spaces for neurodiverse children in the classroom without disrupting classroom flows?*

We try to answer this question following a Design Science Research (DSR) approach [13]. The use of DSR helped us identify design requirements and design objectives relevant to creating in-class calm down spaces that regulate emotional events without making the student feel punished. We began our DSR with a literature review on the design of calm down spaces in educational and psychological literature [14] before creating the prototype. This was followed by a first evaluation workshop with students, who tested and provided feedback on the prototype. Based on this feedback, we aim to conduct at least three additional workshops to test new features with students in addition to expert interviews with education professionals and psychologists.

2 Research Method

We used DSR to inform and evaluate our prototype since DSR is a well-established research method used in the design and development of various socio-technical IT artifacts, such as constructs, frameworks, architectures, models, methods, and instantiations or algorithms [15, 16]. Based on the grounding of our research in neuroscience and psychology, we follow the NeuroIS design science research framework of vom Brocke et al. [17], which includes neurophysiological theories and tools for the building and evaluation phases of the DSR cycle.

Define the Requirements and Objectives of the Solution

Since our goal is to create an affordable, accessible, easy-to-use, virtual environment ("VR app") that acts as an immersive and calming experience, we first derive its requirements and objectives from literature. Research on VR in psychology has already provided valuable insights into anxiety-reducing designs. Valtchanov and Ellard [18], for instance, show that natural environments have a positive effect on anxiety-related emotional events. They use skin-conductance level (SCL) and heart rate variability (HRV) to support the results from two self-report questionnaires on anxiety and affect. Lizsui et al. [19] corroborated their results, adding measures of cortisol concentration. They follow the argumentation that immersion into the VR requires substantial cognitive resources which reduces available capacities for anxiety-related mental processes. Based on this research, we derive two design requirements. *Requirement 1: The design artifact must provide natural environments to have a positive effect on anxiety-related emotional events. Requirement 2: The design artifact must enable a high level of immersion to reduce the cognitive capacity for negative emotional events.*

Research on emotion regulation offers additional insights into the design of a calming VR environment. A prominent process model in this field assumes that every emotion follows an emotion generative process and can be distinguished in antecedent- and response-focused emotion regulation [18]. Antecedent-focused regulation typically emphasizes reappraisal or distraction to break the vicious circle. Response-focused regulation aims to alter and control the response to a completely developed emotion by, for instance, suppression [19]. In this context, breathing exercises are often suggested to achieve emotional regulation and break the vicious circle [20]. Mitsea et al. [21], for instance, show that mindful metacognition in virtual breathing exercises connects attention with respiration, which takes away the attention from the negative stimulus. In a virtual environment, breathing exercises can either be explicit or implicit, with the latter drawing on the effects of mirror neurons [22]. Moreover, emotional regulation is a highly individual process dependent on, for example, personality [22]. Thus, the virtual environment needs to accommodate a high level of individualization to have the desired effect [9]. Based on these findings, we derive two further requirements. *Requirement 3: The design artifact must include attention-commanding elements, such as movement or breathing exercises, to distract from the negative emotions and guide relaxation. Requirement 4: The design artifact must be customizable to allow for effective emotional regulation.*

Since we are proposing the introduction of a piece of technology to the classroom, the final artifact must also meet general objectives of classroom technologies. That is, the VR-application must be (1) quick and easy-to-use, (2) minimally invasive for ongoing lessons, and (3) quickly deployable by the teacher. Most importantly, this solution should (4) achieve a higher order of calming for students experiencing dysregulated emotional events while (5) minimizing disruptions for fellow classmates. Specific objectives regarding the visual (color, saturation, light) and auditory (frequency, harmony, volume) design are derived from, for instance, selective attention studies in autism disorder or ADHD e.g., [23, 24].

Design and Development

Based on our four tentative design requirements and five design objectives, we have started developing a virtual reality-based calm down environment called "Serenity". The prototype application for Serenity is developed in cooperation with Inifa Studios. While the first prototype is already in its final stages, the specific requirements will still be adapted and refined based on student feedback and feedback from professionals. Serenity is being built on a combination of two software platforms—the Google Cardboard Software Development Kit and Unity Software. To allow access to as many teachers as possible, Serenity will be downloadable on IOS and Android devices as a "SmartPhone all" as opposed to a more expensive virtual reality gaming platform such as the Meta Oculus (MSRP $400–500). Once downloaded onto a SmartPhone, the virtual reality aspect of the Serenity app can be experienced using any inexpensive headset viewer, including Google Cardboard (MSRP $10).

Serenity features six unique natural environments which appear to the user as a 360-degree immersive location. Each environment was designed to be visually calming, with further enhancements of relaxing music and sound effects (e.g., the sound of a waterfall). Users (either the student or teacher) select an environment by clicking through a simple menu.

The serenity app allows for the modification of many aspects of the environment, including toggling music and sound effects on and off; the overall color saturation of the scene; and the visual temperature. Modifying the saturation and color temperature of the scenes, in particular, is important to many students on the Autism spectrum susceptible to visual overstimulation [24]. These modifications are accessible via an intuitive dashboard prior to entering each environment. The teacher then places the smartphone into the VR headset and places the headset on the student.

Forthcoming development of the app pertains in particular to Requirement 4. We aim to design an entity, be it a pulsating orb, a favorite animal or cartoon character, or virtual clones, i.e., other students seated nearby in the scene, to guide beathing exercises and to help students regulate anxiety-related negative emotions. We also intend to provide a wider selection of music and sounds. All of these components are under development with the goal of creating a more effective solution than a physical calm down space.

Demonstration

In a first workshop, we conducted semi-structures interviews and performed a small survey with our "proof of concept" prototype at Parson Hills Elementary School, located in Arkansas, USA. In this initial demonstration, a control group of 15 students (3rd–5th grade; ages 9–11) were allowed to experience the Serenity app. For this study, all students were pulled from the general population of the school and were in no way experiencing emotional distress at the time of the study. Each student was allowed to select an environment from the menu, then wear the VR headset and experience the Serenity app for three minutes. Four to five minutes post-experience, students were asked eight questions rating their impression of the app and their overall state of calmness.

We did not yet apply any psychological questionnaires or neurophysiological measures. For the demonstration and evaluation of the next design iteration, we aim to use three psychological questionnaires, Beck Anxiety Inventory [25], Relaxation State Questionnaire [26], and Zuckerman Inventory of Personal Reactions [27]. This way, we hope to capture the level of anxiety before and after the use of our virtual calm down space, and gain insights into the restorative effects of our application. Moreover, we will use neurophysiological measures, such as SCL and HRV to explore the effect of our design. For SCL, we will use the tonic instead of the phasic skin conductance component, since it is associated with tonic states of the sympathetic nervous system arousal, such as stress [28]. Both SCL and HRV will be continuously recoded to measure how long it takes for students to relax with the help of the virtual calm down space. After all, the aim of this application is not to keep students immersed in the VR environment, but to calm them down and get them back into an emotional state that allows them to participate in the classroom.

3 Evaluation

There are several notable outcomes seen in the student response data. For example, when the students were asked if they "felt calmer after using the app", 100% of the students replied affirmatively. When asked if they would like "to use the app again" in their classroom, 100% of the students replied affirmatively. When presented with three options (1: use Serenity in their classroom, 2: use the traditional, physical calm down space; 3: use both in combination) 0% of students wanted to use their class' physical calm down space on its own. 100% of the students selected one of the first two options, both of which included the use of Serenity.

An unanticipated issue arose during the trial which affected 2–3 of the 15 participants. These students expressed a fear that the VR system might contain "scary" content such as "monsters." We attribute this fear to previous experiences with age-inappropriate content on other VR systems or video games, perhaps with older siblings or parents.

Beyond the survey of the 15 "general population" students in the proof-of-concept demonstration, the Infia team was allowed to work with two students identified as being on the Autism spectrum. Our goal was to test whether such students would allow for the physical weight of the VR headset and the strap required to hold the device in place. While initially more anxious than the other students, these students successfully tested the Serenity app and indicated that they had a positive experience. Both students took longer to prove that Serenity would be safe and calming, but once in the environment, neither student wanted to leave.

Acknowledgements This research was funded in part by the Luxembourg National Research Fund (FNR), grant reference 14783405. Moreover, we would like to acknowledge the support of the European Union (EU) within its Horizon 2020 programme, project MDOT (Medical Device Obligations Taskforce), Grant agreement 814654. We also thank the MacMillon Innovation Studio at the University of Arkansas in Fayetteville for their support. For the purpose of open access, the authors have applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

References

1. Chen, Y.-L., & Patten, K. (2021). The Issue Is—Shifting focus from impairments to inclusion: Expanding occupational therapy for neurodiverse students to address school environments. *American Journal of Occupational Therapy, 75*, 7503347010. https://doi.org/10.5014/ajot.2021.040618
2. National Center for Education Statistics. (2023). Back-to-school statistics. https://nces.ed.gov/fastfacts/display.asp?id=372 (last accessed: 12 March 2023).
3. Dieckhaus, M., Hardy, K. K., Anthony, L. G., Verbalis, A., & Pugliese, C. E. (2021). Anxiety relates to classroom executive function problems in students with ASD, but not ADHD. *Research in Autism Spectrum Disorders, 82*(1). https://doi.org/10.1016/j.rasd.2021.101739.

4. Adams, D., McDonald, L., & Keen, D. (2019). Teacher responses to anxiety-related behaviours in students on the autism spectrum. *Research in Developmental Disorders, 86*, 11–19. https://doi.org/10.1016/j.ridd.2018.12.009
5. White, S. W., Oswald, D., Ollendick, T., & Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review, 29*, 216–229.
6. Allen, J. L., & Lerman, R. (2017). Teacher responses to anxiety in children questionnaire (TRAC): Psychometric properties and relationship with teaching staff characteristics. *Emotional and Behavioural Difficulties, 23*(2), 154–168. https://doi.org/10.1080/13632752.2017.1376974. [4] Lantieri, L. (2008). Building inner resilience. *Reclaiming Children and Youth, 17*(2), 43–46.
7. Semple, R., Lee, J., Rosa, D., & Miller, L. (2010). A randomized trial of mindfulness-based cognitive therapy for children: Promoting mindful attention to enhance social-emotional resiliency in children. *Journal of Child and Family Studies, 19*(2), 218–229.
8. Murphey, D., & Sacks, V. (2019). Supporting students with adverse childhood experiences: How Educators and Schools Can Help. *American Educator, 43*(2).
9. Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., & Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review in Psychiatry, 25*, 103–113.
10. Veling, W., Lestestuiver, B., Jongma, M., Hoenders, R., & van Driel, C. (2020). Virtual reality relaxation for patients with a psychiatric disorder: Crossover randomized controlled trial. *Journal of Medical Internet Research, 23*(1), e17233.
11. Wegrzyn, S. C., Hearrington, D., Martin, T., & Randolph, A. B. (2012). Brain games as a potential nonpharmaceutical alternative for the treatment of ADHD. *Journal of Research on Technology in Education, 45*(2), 107–130.
12. Gross, J. J. (1988). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299.
13. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science - Types of knowledge in design science research. *MIS Quarterly, 37*(2), 337–355. https://doi.org/10.2753/MIS0742-1222240302
14. Webster, J., & Watson, R. T. (2002). Analysing the past for prepare the future : Writing a review. *MIS Quarterly, 26*(2).
15. Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*.
16. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302
17. Vom Brocke, J., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems, 53*(3), 1–13. https://doi.org/10.1080/08874417.2013.11645627
18. Valtchanov, D., & Ellard, C. (2010). Physiological and affective responses to immersion in virtual reality: Effects of nature and urban settings. *Journal of Cybertherapy and Rehabilitation, 3*, 359–374.
19. Liszio, S., Graf, L., & Masuch, M. (2018). The relaxing effect of virtual nature: Immersive technology provides relief in acute stress situations [Article]. *Annual Review of Cybertherapy and Telemedicine, 2018*(16), 87–93. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067858592&partnerID=40&md5=1fd197a6062eb9de428267c3cb6fe659.
20. Astor, P. J., Adam, M. T. P., Jerčic, P., Schaaff, K., & Weinhardt, C. (2013). Integrating biosignals into information systems: A neurois tool for improving emotion regulation. *Journal of Management Information Systems, 30*(3), 247–278.
21. Mitsea, E., Athanasios, D., & Skianis, C. (2022). Breathing attention & consciousness in sync: The role of breathing training, metacognition & virtual reality. *Technicum Social Sicences Journal, 29*, 79–97.
22. Shafir, T. (2016). Using movement to regulate emotion: Neurophysiological findings and their application in psychotherapy. *Frontiers in Psychology, 7*, 14–51. https://doi.org/10.3389/fpsyg.2016.01451

23. Tyndall, I., Ragless, L., & O'Hora, D. (2018). Effects of perceptual load and socially meaningful stimuli on crossmodal selective attention in Autism Spectrum Disorder and neurotypical samples. *Conciousness and Cognition, 60*, 25–36. https://doi.org/10.1016/j.concog.2018.02.006
24. Dursun, P., Fidan, U., & Karayagiz, S. (2021). Probable role of listening therapy in the management of ADHD symptoms: Three case studies. *Current Psychology, 40*, 4219–4234. https://doi.org/10.1007/s12144-021-01419-x
25. Beck, A. T. & Steer, R. A. (1993). *Beck anxiety inventory manual*. Psychological Corporation.
26. Steghaus, S., & Poth, C. H. (2022). Assessing momentary relaxation using the Relaxation State Questionnaire (RSQ). *Science and Reports, 12*, 16341. https://doi.org/10.1038/s41598-022-20524-w
27. Zuckerman, M. (1977). Development of a situation-specific trait-state test for prediction and measurement of affective responses. *Journal of Consulting and Clinical Psychology, 45*(4), 513–523.
28. Valtchanov, D., Barton, K.R., & Ellard, C. (2010). Restorative effects of virtual nature settings. *Cyberpsychology, Behavior, and Social Networking, 13*(5), 503–512. https://doi.org/10.1089/cyber.2009.0308

Author Index

F. D. Davis et al. (eds.), *Information Systems and Neuroscience*, Lecture Notes in Information Systems and Organisation 68, https://doi.org/10.1007/978-3-031-58396-4