Proceedings
NeuroIS Retreat 2026

Vienna, Austria | June 2-4
www.NeuroIS.org

Fred D. Davis, René Riedl,
Jan vom Brocke, Pierre-Majorique Léger,
Adriane B. Randolph, Gernot R. Müller-Putz
(Eds.)

The final proceedings will be published by Springer.

Preface

The proceedings contain papers presented at the 18th annual NeuroIS Retreat held June 2-4, 2026. 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 2010-2017. Beginning in 2018, the conference is taking place in Vienna, Austria.

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 eleventh 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 2026 keynote presentation entitled "Show Me Your Smartphone and Then I Will Show You Your Neurobiology" was given by Christian Montag, Distinguished Professor for Cognitive and Brain Sciences, and Associate Director of the Institute of Collaborative Innovation at the University of Macau, China. Moreover, Frank Krueger, Professor of Systems Social Neuroscience at George Mason University, United States, gave a hot topic talk entitled "Neurophysiological Dynamics of Trust in Human–AI Interaction: A Multi-Level Study of Brain, Hormone, Mind, and Behavior".

Altogether, we are happy to see the ongoing progress in the NeuroIS field. We are also pleased to report that the NeuroIS Society, founded in 2018 as a non-profit organization, has been progressing positively. We anticipate continued and prosperous growth in the field of NeuroIS.

June 2026

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

Sponsors

We thank the sponsors of the NeuroIS Retreat 2026.

Main Sponsors

Technology Sponsors

Keynote

Show Me Your Smartphone and Then I Will Show You Your Neurobiology

Christian Montag

Digital phenotyping describes the endeavor to study digital footprints from various sources such as smartphones or social media to predict various psychological traits and states. Only few studies went one step further, namely aiming at the prediction of neurobiological variables from digital traces. In the presentation an overview will be provided on current advances in the field by also reflecting on how artificial intelligence can help to make accurate predictions to come up with what is called "digital biomarkers". In this context also AI chatbots – often used via the smartphone - will be discussed as an additional data source to carry out digital phenotyping procedures. Technical and ethical challenges will be another central focus of the presentation.

Hot Topic Talk

Neurophysiological Dynamics of Trust in Human–AI Interaction: A Multi-Level Study of Brain, Hormone, Mind, and Behavior

Frank Krueger

As artificial intelligence (AI) increasingly participates in human decision environments, understanding how trust emerges and breaks down in human–AI interaction has become a central challenge for information systems research. While prior work has predominantly focused on behavioral measures, the underlying neurophysiological mechanisms of trust remain largely unexplored. In this study, we investigated the multi-level dynamics of trust during face-to-face interaction with an embodied intelligent agent. Participants engaged in decision-making tasks with a humanoid robot while neural activity was recorded using functional near-infrared spectroscopy, salivary oxytocin levels were assessed, and self-reported trust and behavioral influence were measured. We experimentally manipulated system reliability (congruent vs. erroneous decisions) and social expressiveness (animated vs. stationary behavior). Results demonstrated that reliability constitutes the primary foundation of trust: agent errors significantly reduced both reported trust and behavioral influence. Social expressiveness modulated these effects. Animated agents elicited stronger prefrontal activation and enhanced neural–hormonal coupling. Notably, elevated oxytocin levels were associated with reduced trust and diminished behavioral influence when expressive agents committed errors, indicating a context-sensitive vigilance response rather than a simple affiliative bonding mechanism. Together, these findings establish a multi-level neurophysiological framework for understanding trust in human–AI interaction and reveal a critical design trade-off between social expressiveness and trust robustness in intelligent systems.

Table of Contents

Assessing Task Load using Functional Near-Infrared Spectroscopy

Alexander John Karran[1*], Thaddé Rolon-Merette[1], Kellyann Stamp[2], and Stephen Fairclough[3], Sylvain Sénécal[1], Pierre-Majorique Léger[1], and Constantinos K. Coursaris[1]

[1] Department of Information Technologies, HEC Montréal, Montréal, Québec, Canada
{alexander-john.karran, thadde.rolon-merette, sylvain.senecal,pierre-majorique.leger, constantinos.coursaris}@hec.ca

[2]School of Computer Science and Mathematics, [3]School of Natural Sciences and Psychology, Liverpool John Moores, Liverpool United Kingdom
{K.A.Stamp, s.fairclough}@ljmu.ac.uk

Abstract. Measuring task load (TL) in complex environments is critical for optimizing human-computer interaction. While functional near-infrared spectroscopy (fNIRS) offers objective, real-time TL measurement, consensus on optimal cortical markers remains elusive. This study investigated multiregional fNIRS hemodynamics and subjective NASA-TLX scores in 42 adults during a simulated driving task across three difficulty levels. Results showed that subjective TL increased linearly with difficulty. However, fNIRS revealed diverse regional dynamics: frontopolar activation increased linearly, left dorsolateral prefrontal cortex activation decreased, and right premotor regions exhibited non-linear (U-shaped) responses. These findings highlight TL's distributed nature, emphasizing the need for multi-regional measures of TL.

Keywords: Task Load, fNIRS, Driving, Human-Computer Interaction

Introduction

Measuring cognitive effort is a crucial construct in Information Systems (IS), psychology, and human-computer interaction research [1-3]. Interacting with complex technology places significant demands on users' attentional and executive resources [4]. Dynamic fluctuations in task demand cause concomitant variations in cognitive effort, which can degrade the quality of human-machine interaction. Such variations may lead to disengagement in cases of low cognitive demand, or impaired decision-making, increased errors, and diminished system usability [5] in cases of excessive cognitive demand. Thus, an integral goal of system design should be to maintain an optimal level of mental effort. Task load (TL) is conceptualized here as a broader operational construct than either cognitive or mental load; it integrates the cognitive, psychomotor, and situational demands that emerge during continuous interaction with complex systems. Cognitive load [4] narrows this to information-processing demand on working memory, whereas mental load [6] is closer to TL but centers on attentional resource allocation. Neuroadaptive systems research has demonstrated that maintaining optimal TL can enhance task performance or increase safety [7-9]. In ecologically valid contexts such as driving or flight, these demands may not be reflected by a single cortical marker or by retrospective self-reports alone. Researchers often depend on subjective self-reports, such as the NASA Task Load Index (NASA-TLX) [10], which treats workload as the interaction between task demands and operator resources, spanning the mental, physical, and temporal dimensions, as well as performance, effort, and frustration. While this instrument is widely used to identify workload sources, it is a retrospective, perceptual measure that is unable to capture dynamic, real-time fluctuations [11]. Given that users process multidimensional information in real-world scenarios, a continuous, objective measurement of TL is the most ecologically valid target for information systems seeking to adapt to operational reality [12].

This highlights the importance of NeuroIS approaches that enable systems to track ongoing effort using objective physiological measures such as pupillometry, electrocardiogram (ECG), electroencephalography (EEG) and fNIRS [13-15]. However, EEG's susceptibility to movement artifacts limits its utility in mobile, naturalistic settings. Consequently, functional near-infrared spectroscopy (fNIRS) is often preferred in dynamic contexts where mobility is an asset. fNIRS is a portable optical neuroimaging technique that uses the modified Beer–Lambert Law to measure localized cortical hemodynamics, specifically changes in oxygenated (HbO) and deoxygenated (HbR) hemoglobin commonly referred to as the blood oxygen level dependent response (BOLD) [16]. It offers superior spatial resolution and resistance to motion artifacts compared to EEG, making it highly suitable for ecologically valid environments, albeit with lower temporal resolution [17-19].

Despite these advantages, there is a lack of consensus on the optimal fNIRS markers of TL in ecologically valid environments. Furthermore, the literature yields conflicting results regarding the relationship between HbO and HbR and task difficulty. Studies using abstract working memory paradigms typically observe a linear increase in prefrontal cortex (PFC) oxygenation as difficulty rises [20, 21]. Conversely, studies using simulated piloting or driving often report a non-linear pattern, such as a U or inverted U-shape. For instance, this could be when PFC activation drops when task demands exceed mental capacity, signalling overload, and a withdrawal of cognitive resources. Or a low PFC activation when the task is not engaging enough [22-26]. Most of these studies examine the PFC exclusively for measures of task-related mental effort. This narrow focus often overlooks distinct cortical regions potentially involved in TL processing, such as the anterior prefrontal cortex or frontopolar area (FPA), dorsolateral (dlPFC), and dorsal pre-motor (PMd) [27].

As such, the convergence of objective fNIRS signals across various cortical regions remains underexplored. Therefore, the present study investigated fNIRS-based measurement of task load during a simulated driving task with parametrically manipulated difficulty levels (easy, medium, and hard). A driving task served as an ecologically grounded context in which task load arises from visuospatial processing, motor control, and temporal pressure. Continuous fNIRS data were acquired from 38 healthy adults over frontal and premotor regions, alongside subjective NASA-TLX assessments. Using a general linear model (GLM) framework, we analyzed channel-wise hemodynamic responses to characterize the spatial profile of cortical activation across difficulty levels and to evaluate whether objective fNIRS signals consistently track multidimensional task load alongside subjective ratings. We therefore ask whether prefrontal and premotor regions exhibit dissociable hemodynamic responses to parametrically increased difficulty in a continuous simulated driving task, and whether those regional responses align with, or diverge from, subjective NASA-TLX ratings. We hypothesized that subjective task load would scale monotonically with difficulty, while fNIRS would reveal region-specific dynamics, some linear, some non-linear consistent with a distributed rather than unitary cortical representation of task load.

Methods

Experimental design

Forty-two healthy adults (20 Female, aged 18-40, average 26.7 years) were recruited via university advertisements. Participants provided written informed consent prior to participation and were compensated with CAD $40 upon completion of the experiment. Participants performed six driving game tasks under 3 conditions, derived from [28], consisting of three 4-min blocks at “easy”, “medium”, and “hard” difficulty levels. The tasks (see figure 1) required participants to drive a vehicle in the “wrong direction” i.e., against oncoming traffic on a four-lane road populated with other vehicles. The goal was to avoid collisions with oncoming vehicles by maneuvering laterally from left to right. Game difficulty was manipulated by increasing the speed of the participant's vehicle; thus, participants had less time at high speed to respond to oncoming vehicles and avoid collisions. Players accumulated a total game score by avoiding collisions; points accrue with every passing second. However, a penalty was deducted from this score upon a collision, reducing the total game score. The task order was counterbalanced to correct for carryover effects. Data collection occurred in an acoustically shielded testing room. The game was launched in windowed mode (1440×900, Ultra graphics). Two Alienware PCs were used: one running the driving game and one controlling the fNIRS acquisition via NIRS STAR (NIRx Medical Technologies). Light was controlled to minimize ambient interference. Immediately after completing each individual driving trial, the simulation paused, and participants completed the NASA-TLX presented via iPad to reflect on the trial they had just finished.

a) b)

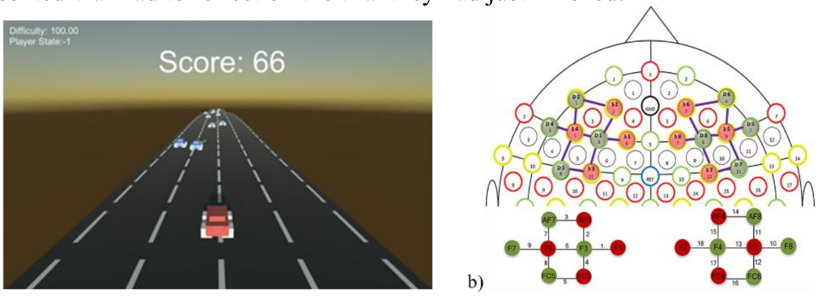


Fig. 1. The (a) driving game and (b) fNIRS montage.

Task load and analysis

Subjective TL was assessed using an average of six NASA-TLX subscales for each trial. The effect of task difficulty on subjective TL was tested using a repeated-measure ANOVA, followed by Bonferroni-correct pairwise comparisons. Objective TL was assessed using fNIRS and a GLM analysis of HbO and HbR activity across predefined regions of interest: specifically, the frontopolar area (FPA, *BA 10*), left and right dorsolateral prefrontal cortex (DLPFC, *BA 9/46*), and left and right dorsal premotor cortex (PMd, *BA 6*). fNIRS data were acquired at 7.8 Hz using a NIRSport system (NIRx Medical Technologies) with dual wavelengths (760 nm and 850 nm). An 8 × 8 optode montage yielding 24 measurement channels was positioned on a 58 cm cap and aligned to Cz using standard nasion-inion reference measurements (see Figure 1b). Table 1 summarizes the mapping between the predefined fNIRS ROIs, their constituent source-detector channels, and the corresponding approximate 10-20 scalp landmarks.

Table 1. Mapping of fNIRS regions of interest (ROI) to source-detector channels and approximate 10-20 landmarks

ROI	Label	Channels	Approx. 10-20 landmarks
Left dlPFC	L-DLPFC	S2_D2, S2_D4, S4_D4	AF3, F3
Right dlPFC	R-DLPFC	S5_D5, S6_D5, S6_D6	AF4, F4
Frontal Pole Area	FPA	S1_D1, S2_D2, S8_D1, S1_D8, S6_D6, S8_D8	AFz, Fz
Left VLPFC	L-VLPFC	S3_D4, S4_D3	F5, F7
Right VLPFC	R-VLPFC	S7_D5, S5_D7	F6, F8
Left Premotor/SMA	L-PMd	S3_D1, S3_D3, S4_D1	FC3, FC5
Right Premotor/SMA	R-PMd	S7_D7, S7_D8, S5_D8	FC4, FC6

Signal quality was calibrated using NIRS STAR software prior to task onset, and at least 95% of channels met optimal coupling thresholds. Data preprocessing was conducted using MNE-NIRS (v. 0.71). Raw light intensities were converted to optical density and then transformed into concentration changes of oxyhemoglobin (ΔHbO) and deoxyhemoglobin (ΔHbR) using the modified Beer-Lambert Law [16] with a differential pathlength factor of 6. Signal quality was evaluated using the scalp coupling index (SCI); channels with an SCI < 0.75 were excluded, resulting in the rejection of 5.7% of total data. Each of the three 4-min difficulty blocks (easy, medium, hard) was entered as a separate event marker modeled as a boxcar of 240s duration, and the hemodynamic response was estimated with a finite impulse response (FIR) basis spanning 0-15s post-onset, sampled at the 7.8 Hz acquisition rate. The 15-s epochs thus index the within-block response kernel recovered by the FIR, not a per-trial segmentation. For each subject, a first-level design matrix was generated with a third-order polynomial drift model and a finite impulse response haemodynamic model. The GLM was fit channel-wise to HbO and HbR time courses using ordinary least squares, yielding β-coefficients for each condition. At the group level, per-participant first-level condition β-coefficients were carried forward, and β-coefficients were aggregated across participants and contrasted pairwise via two-tailed paired *t*-tests on the three condition contrasts (easy vs. medium, easy vs. hard, medium vs. hard), computed separately for HbO and HbR at each ROI. All statistical tests were Bonferroni corrected across the family of pairwise contrasts within each ROI × condition combination where $p < 0.05$ were deemed significant.

Results

Subjective TL varied significantly as a function of task difficulty, with NASA-TLX scores increasing across the easy, medium, and hard conditions ($F(2, 84) = 37.70$, $\eta^2_p = .47$, $p < .001$). Bonferroni-corrected pairwise comparisons showed that the easy (M = 51.79, SD = 13.14) condition yielded lower TL than the medium and hard conditions, and the medium condition was lower than the hard condition.
The repeated measures ANOVA on subjective TL scores indicated a significant main effect of task difficulty on subjective workload scores, $F(2, 84) = 37.70$, $p < .001$, with scores increasing linearly with task difficulty. They were significantly lower in the Easy condition (M = 51.79, SD = 13.14) than the Medium (M = 56.31, SD = 13.12, $t(42) = -4.23$, $d_z = 0.65$, $p < .001$) and the Hard (M = 61.89, SD = 13.36, $t(42) = -6.79$, $d_z = 1.04$, $d_z = 1.01$, $p < .001$) as well as lower in the Easy than in the Hard condition ($t(42) = -6.61$, $p < .001$).

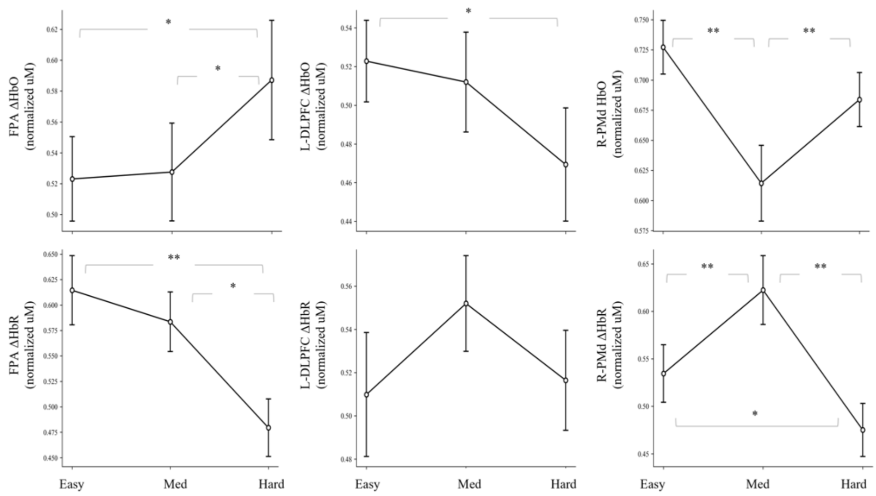


Fig. 2. Results from fNIRS GLM. ΔHbO/ΔHbR acrss tasks and ROIs. * $p < .05$; ** $p < .01$.

For objective TL, significant differences were found at the FPA, left dlPFC and right PMd regions, as shown in Figure 2. For FPA, a significant effect of difficulty was observed for HbO ($F(2, 74) = 3.46$, $p = .037$). Pairwise comparisons showed that differences followed a linear increase pattern, with the lowest activation in the Easy condition compared to the highest Hard condition ($t = 2.16$, $p = .039$). Medium showed a higher descriptive HbO level than Easy but significantly lower than Hard (t = -2.09, p = .044). Conversely, HbR showed a significant main effect, $F(2, 74) = 6.80$, p = .01, and pairwise tests revealed a linear decrease, with Easy (t = 2.96, p = .007) and Medium (t = 2.10, p = .044) higher than Hard. In the left dlPFC, a linear decrease trend was found in HbO ($F(2, 74) = 2.58$, p = .082), with Easy higher than Hard (t = 2.76, p = .011). No significant main or pairwise effect for HbR was reported. No significant effects were reported in the right dlPFC. However, a significant main effect for HbO was observed at the right PMd region ($F(2, 74) = 5.15$, p = .008), which displayed a U-shaped trend as the medium condition showed a lower level than Easy (t = 3.19, p = .006) and Hard (t = 3.17, p = .008). A significant main effect of condition was reported in the right PMd for HbR ($F(2, 74) = 5.11$, p = .007), pairwise tests showed a higher response in the Medium condition than Easy (t = 2.99, p = .009) and Hard (t = 3.11, p = .008) and Easy higher than Hard (t = 2.86, p = .017). No significant differences were found in the left PMd region.

Discussion

This study aimed to examine cortical regions related to TL during a continuous simulated driving task across three difficulty levels. While subjective TL increased from easy to hard, objective TL, captured by ROIs GLM analysis of several ROIs using fNIRS, TL revealed both linear and non-linear patterns in different ROIs. While these were broadly congruent with the subjective TL, they highlighted the distributed nature of TL across task difficulty levels. Specifically, the FPA region exhibited a positive linear relationship between HbO and task difficulty, and a linear decrease in HbR. Conversely, the left dlPFC displayed a negative linear trend in HbO, with activation decreasing with difficulty. These results align with previous fMRI results, indicating positive linear increases in brain activity in frontal areas and a decrease in medial and dlPFC areas [20, 29, 30]. FPA results also aligned with previous fNIRs studies and support that frontal areas activate in response to working memory and executive demands [31], while DLPFC regions could deactivate with task difficulty to inhibit states such as mind-wandering [32]. Furthermore, the right PMd demonstrated non-linear dynamics, characterized by a U-shaped HbO response and an inverted U-shaped HbR response. These non-linear patterns align with results from previous fNIRS studies [26] and behavioural preferences during driving tasks [33]. We note, however, that three difficulty levels constitute the minimum number of points required to detect departures from monotonicity; the detected U-shaped pattern in right PMd HbO is consistent with non-linear load-dependent dynamics. However, future work that replicates this result with a finer parametric gradation is required to verify this pattern of activity.

Interestingly, HbO and HbR often exhibited opposite patterns within the same cortical region. This neurovascular decoupling does not align with the expected coupled hemodynamic response function (HRF); it may be that the continuous, temporal nature of the simulated driving task led to a deviation from the standard model. Under heavy load, the neurovascular response essentially decouples, favouring HbR. Potentially due to increased metabolic processes in which the brain must continually flush out accumulated reactive oxygen species and other by-products to sustain effortful task completion, thus, HbR dynamics become increasingly prominent. An alternative interpretation for the observed neurovascular decoupling effect could be due to non-region specific haemodynamics, in that HbO and HbR activity in the FPA (inverse) and right PMd (expected) taken together represent the canonical signature of a well coupled cortical haemodynamic response [15]. In this view regions showing this coupled pattern across conditions can therefore be interpreted as genuine task-load-sensitive activations rather than artifacts of motion, systemic physiology, or extracerebral scalp signal. Furthermore, this interpretation leads to a conclusion that the divergent regional activity: linear in FPA, inverted in L-DLPFC, U-Shaped in R-PMd are not inconsistencies but potential evidence that the signal subjective rating provided by the NASA-TLX is collapsing three dissociable cortical dynamic effects into one number.

Future analysis could investigate this by examining functional connectivity among the dlPFC, FPA, PFC, and PMd regions and how deeper neural structures interact to manage task load. Furthermore, the potential for multi-regional hemodynamic variation suggests that traditional linear models cannot fully capture these complexities. Consequently, future research could also use a Deep Learning approach and perhaps develop a highly accurate, real-time fNIRS task load index [34-37]. This could provide groundwork for neuroadaptive systems that aim to adapt interface complexity in response to ongoing cortical dynamics, with the longer-term goal of supporting engagement and safer interactions in ecologically valid settings.

References

1. Blankertz, B., et al., *The Berlin brain–computer interface: non-medical uses of BCI technology.* Frontiers in neuroscience, 2010. **4**: p. 2050.
2. Cinel, C., D. Valeriani, and R. Poli, *Neurotechnologies for human cognitive augmentation: current state of the art and future prospects.* Frontiers in human neuroscience, 2019. **13**: p. 13.
3. Marblestone, A.H., G. Wayne, and K.P. Kording, *Toward an integration of deep learning and neuroscience.* Frontiers in computational neuroscience, 2016. **10**: p. 94.
4. Sweller, J., *Element interactivity and intrinsic, extraneous, and germane cognitive load.* Educational psychology review, 2010. **22**: p. 123-138.
5. Paas, F., A. Renkl, and J. Sweller, *Cognitive load theory and instructional design: Recent developments.* Educational psychologist, 2003. **38**(1): p. 1-4.
6. Wickens, C.D., *Multiple resources and mental workload.* Human factors, 2008. **50**(3): p. 449-455.
7. Fairclough, S., *Neuroadaptive technology and the self: A postphenomenological perspective.* Philosophy & Technology, 2023. **36**(2): p. 30.
8. Krol, L.R., P. Haselager, and T.O. Zander, *Cognitive and affective probing: a tutorial and review of active learning for neuroadaptive technology.* Journal of neural engineering, 2020. **17**(1): p. 012001.
9. Mark, J.A., et al., *Neuroadaptive training via fNIRS in flight simulators.* Frontiers in Neuroergonomics, 2022. **3**: p. 820523.
10. Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*, in *Advances in psychology*. 1988, Elsevier. p. 139-183.

11. Hart, S.G. *NASA-task load index (NASA-TLX); 20 years later*. in *Proceedings of the human factors and ergonomics society annual meeting*. 2006. Sage publications Sage CA: Los Angeles, CA.
12. Parasuraman, R., *Neuroergonomics: Brain, cognition, and performance at work.* Current directions in psychological science, 2011. **20**(3): p. 181-186.
13. Das Chakladar, D. and P.P. Roy, *Cognitive workload estimation using physiological measures: a review.* Cognitive Neurodynamics, 2024. **18**(4): p. 1445-1465.
14. Dimoka, A., et al., *On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for Neurois1.* MIS quarterly, 2012. **36**(3): p. 679-702.
15. Quaresima, V. and M. Ferrari, *Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review.* Organizational Research Methods, 2019. **22**(1): p. 46-68.
16. Huppert, T.J., et al., *HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain.* Applied optics, 2009. **48**(10): p. D280-D298.
17. Martini, M.L., et al., *Sensor modalities for brain-computer interface technology: a comprehensive literature review.* Neurosurgery, 2020. **86**(2): p. E108-E117.
18. Naseer, N. and K.-S. Hong, *fNIRS-based brain-computer interfaces: a review.* Frontiers in human neuroscience, 2015. **9**: p. 3.
19. Scholkmann, F., et al., *A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology.* Neuroimage, 2014. **85**: p. 6-27.
20. Causse, M., et al., *Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS.* Scientific reports, 2017. **7**(1): p. 5222.
21. Herff, C., et al., *Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS.* Frontiers in human neuroscience, 2014. **7**: p. 935.
22. Ayaz, H., et al., *Optical brain monitoring for operator training and mental workload assessment.* Neuroimage, 2012. **59**(1): p. 36-47.
23. Gateau, T., et al., *Real-time state estimation in a flight simulator using fNIRS.* PloS one, 2015. **10**(3): p. e0121279.
24. Yoshino, K., et al., *Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway.* Frontiers in human neuroscience, 2013. **7**: p. 882.
25. Herff, C., F. Putze, and T. Schultz. *Evaluating fNIRS-based workload discrimination in a realistic driving scenario*. in *The First Biannual Neuroadaptive Technology Conference*. 2017.
26. Durantin, G., et al., *Using near infrared spectroscopy and heart rate variability to detect mental overload.* Behavioural brain research, 2014. **259**: p. 16-23.
27. Gemmerich, R., O. Müller, and A. Schaller, *The application of fNIRS in studies on occupational workload: a systematic review.* Frontiers in Public Health, 2025. **13**: p. 1560605.
28. Fairclough, S.H., K. Stamp, and C. Dobbins, *Functional connectivity across dorsal and ventral attention networks in response to task difficulty and experimental pain.* Neuroscience Letters, 2023. **793**: p. 136967.
29. Arsalidou, M., et al., *A balancing act of the brain: Activations and deactivations driven by cognitive load.* Brain and behavior, 2013. **3**(3): p. 273-285.

30. Yu, K., et al., *Exploring cognitive load through neuropsychological features: an analysis using fNIRS-eye tracking.* Medical & Biological Engineering & Computing, 2025. **63**(1): p. 45-57.
31. Midha, S., et al., *Measuring mental workload variations in office work tasks using fNIRS.* International Journal of Human-Computer Studies, 2021. **147**: p. 102580.
32. Gu, H., et al., *Regional excitation-inhibition balance predicts default-mode network deactivation via functional connectivity.* Neuroimage, 2019. **185**: p. 388-397.
33. Ma, S., et al., *Promote or inhibit: An inverted U-shaped effect of workload on driver takeover performance.* Traffic injury prevention, 2020. **21**(7): p. 482-487.
34. Chiarelli, A.M., et al., *Deep learning for hybrid EEG-fNIRS brain–computer interface: application to motor imagery classification.* Journal of neural engineering, 2018. **15**(3): p. 036028.
35. Eastmond, C., et al., *Deep learning in fNIRS: a review.* Neurophotonics, 2022. **9**(4): p. 041411-041411.
36. Khan, M.A., et al., *Application of artificial intelligence in cognitive load analysis using functional near-infrared spectroscopy: A systematic review.* Expert systems with applications, 2024. **249**: p. 123717.
37. Park, J.-H., *Mental workload classification using convolutional neural networks based on fNIRS-derived prefrontal activity.* BMC neurology, 2023. **23**(1): p. 442.

The Influence of Breathing on Brain and Heart Activity during Motor Execution

Laura Maria Höber[1], Kyriaki Kostoglou[1], Gernot R. Müller-Putz[1,2,*]

[1] Graz University of Technology, Institute of Neural Engineering, Graz, Austria
[2] BiotechMed, Graz, Austria
gernot.mueller@tugraz.at

Abstract. Electroencephalography (EEG) is widely used in brain–computer interfaces (BCIs) to detect voluntary movement initiation, particularly in motor-impaired users. Yet movement intention is typically studied in isolation, despite emerging within ongoing cardiorespiratory dynamics. We examined the coupling of brain, respiratory, and cardiac activity during self-paced and breathing-cued hand movements. Simultaneous EEG, electrocardiography (ECG), and respiration recordings showed that paced breathing enhanced movement-related EEG responses and heart rate modulation, whereas natural breathing produced subtler effects, with movements preferentially initiated during expiration. These findings indicate that cardiorespiratory state shapes cortical markers of movement intention, with implications for improving BCI detection and, more broadly, enhancing robustness and personalization in real-time human–technology interaction.

Keywords: MRCP · self-paced movement · cardiorespiratory coupling · BCI · multimodal biosignals

Introduction

Detecting voluntary movement intentions from electroencephalography (EEG) is central to non-invasive brain–computer interfaces (BCIs), where attempted or imagined movements are translated into discrete control signals [1]. Movement intention is typically reflected in the movement-related cortical potential (MRCP), a slow negative potential emerging up to two seconds before and peaking at movement onset [2,3]. However, MRCPs are weak, especially in motor-impaired users who can only attempt or imagine movements, which complicates reliable intention decoding from brain signals alone.

Voluntary actions, however, arise not only from neural processes but are embedded in ongoing physiological rhythms such as breathing and cardiac activity [4,5]. Respiration is especially relevant as a dominant rhythm under voluntary and autonomic control, generated in the brainstem and modulated by cortical and emotional influences [6,7,8]. Beyond respiratory motor control, breathing modulates neural excitability and biases the brain toward exteroceptive or interoceptive processing [9]. This affects stimulus uptake and reaction timing [10,11,12], with voluntary movements preferred during expiration and transitions between phases largely avoided [13,14,15,16,17]. Furthermore, the MRCP amplitude during voluntary movement generation is modulated by breathing phase, with greater negativity during inspiration [18], [19].

Cardiac activity is coupled to respiration via respiratory sinus arrhythmia (RSA), with heart rate (HR) increasing during inspiration and decreasing during expiration [20,21]. Voluntarily initiated movements also induce transient cardiac responses [22], with a decrease in HR before movement onset and an increase thereafter. How these RSA and movement-related effects interact remains unclear.

In this study, healthy participants serve as a model system to investigate integrated physiological processes underlying voluntary movement initiation. We simultaneously examine EEG, respiration, and cardiac activity during self-paced and breathing-cued hand movements, addressing three research questions: (RQ1) which breathing phase favors movement onset, (RQ2) how breathing-phase alignment affects HR dynamics, and (RQ3) how it influences MRCPs. By integrating neural, respiratory, and cardiac signals, this work moves beyond isolated MRCP analysis and toward a more comprehensive characterization of the physiological state preceding voluntary action. These insights are relevant not only for improving movement intention decoding in BCIs, but also for neuroinformation systems (NeuroIS), where multimodal physiological integration can enhance adaptive system design, user-state modeling, and the reliability of real-time human–technology interaction.

Methods and Materials

Participants and Experimental Setup

Twenty healthy participants (6 female, 26.1±2.8 years) took part in the study; nineteen were right-handed according to the Edinburgh Handedness Inventory [23]. The single left-handed participant was not analyzed separately, as no systematic differences between dominant and non-dominant hand movements have been reported [24].

EEG was recorded using a 64-channel eego-sports system (ANT Neuro, Hengelo, Netherlands) with ground at CPz and reference at Fpz. Respiration (elastic chest belt, SleepSense, Tel Aviv, Israel), EMG (bipolar electrodes over the right extensor carpi radialis and ulnaris; ground on the left styloid), and ECG (Einthoven II derivation) were recorded using a g.USBAMP amplifier (g.tec medical engineering GmbH, Schiedelberg, Austria). All signals were sampled at 512 Hz and synchronized via Lab Streaming Layer (LSL) [25]. The recordings took place in a shielded room, with participants sitting in front of a screen and keyboard, and both hands resting on the table in front of them.

The experiment began with 2 min of eyes-open rest. Participants then performed three 200-s runs of self-paced right-wrist flexions, separated by 30-s breaks and preceded by a 1-min training run. They were instructed to initiate movements voluntarily, maintaining an approximate 10-s interval between movements without explicitly counting, to discourage rhythmic or automatic execution. After a 5-min break, two cued conditions followed, with movements triggered at the end of inspiration or expiration, respectively. Each condition included three 200-s runs with 30-s breaks and a 1-min training run in pseudo-randomized order. Each cued trial started with a random pause of

5.5±1 s, followed by visually guided breathing, and movement cues given at 66% of the respective last breathing phase. The guidance was given by a blue circle in the middle of the fixation cross, expanding over 1.65 s for inspiration, contracting over 2 s for expiration, and crossing a transparent line to cue the movement (Fig. 1).

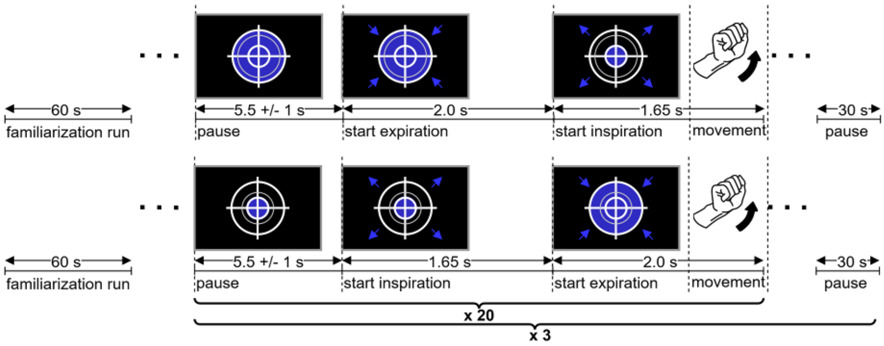


Fig. 1: Sequence of the movement paradigm under cued breathing conditions. The circle's expansion guided inspiration, its contraction guided expiration, and the crossing of a slightly transparent line instructed the start of a movement.

Signal Preprocessing and Statistical Analysis

All signals were band-pass filtered using a 4th-order Butterworth filter (band-specific ranges indicated below), with an additional 50 Hz notch filter to remove power-line interference. EMG (20–200 Hz) was rectified, smoothed (0.2-s window), and movement onset was defined as the first zero-crossing of the EMG derivative preceding the main peak. The respiration signal (0.1–2 Hz) was smoothed (2-s window), and breathing-phase transitions were identified from derivative zero-crossings, excluding intervals shorter than 0.8 s. ECG (0.3–30 Hz) R-peaks were detected using the Pan–Tompkins algorithm [26]. Instantaneous HR was derived from the R–R intervals, and HR variability (HRV) spectral power was estimated using Welch's method [27,28] in the very low (VLF), low (LF), and high frequencies (HF). EEG was downsampled to 128 Hz, band-pass filtered (0.1–64 Hz), and cleaned from artefactual components related to eye movements, muscle activity, and channel noise using independent component analysis (ICA) in EEGLab [29]. To isolate the low-frequency MRCPs, an additional low-pass filter with a cutoff frequency of 3.5 Hz was applied. Finally, the signals were re-referenced to the common average to remove spatially widespread artefacts, including those potentially arising from breathing-related movements. All signals were then epoched around movement onset, except for the rest condition, where windows were centered around the start of expiration. Trials with noisy signals or incorrect breathing-phase alignment were excluded.

Pairwise t-tests with Benjamini–Hochberg FDR correction [30] were used to compare self-paced and cued movements during inspiration and expiration; resting data were included when appropriate. The distribution of movement onsets within the breathing cycle was tested for non-uniformity using an r-test [31], and correlations between respiratory, cardiac, and EEG-derived measures were analyzed per condition. All processing and subsequent visualizations were implemented in MATLAB [32].

Results

RQ1: We examined respiratory characteristics and the distribution of movement onsets across the breathing cycle. Breathing rate was highest during self-paced trials (0.27 Hz) and lower during cued breathing, which was comparable to rest (0.22 Hz). Movement onsets showed a significant preference for early to mid-expiration, while phase transitions were largely avoided (Fig. 2a). Breathing amplitude (Fig. 2b) was shallower during self-paced movements and larger during rest and cued breathing.

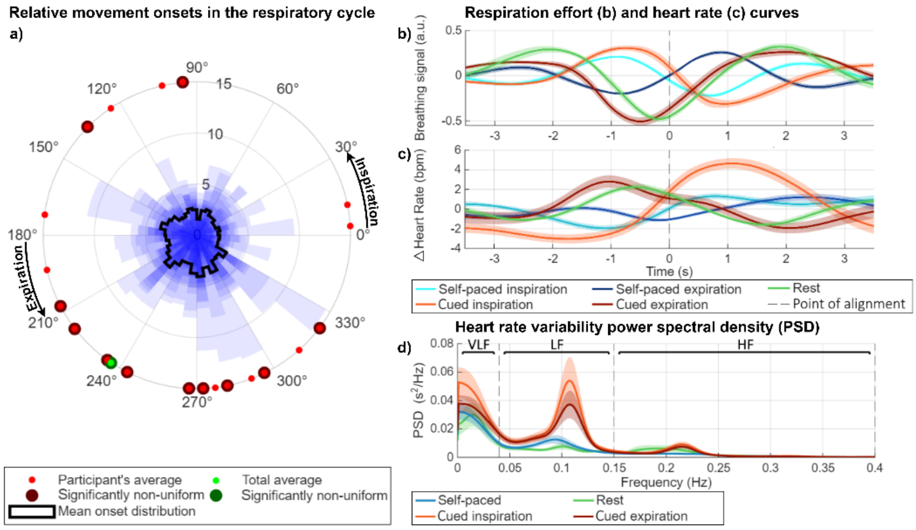


Fig. 2: (a) Individual distributions of relative movement onsets across the respiratory cycle (blue histograms) and grand average distribution (black outline). Individual mean onsets are shown as red dots, and the grand average onset as a green dot. Mean onsets from significantly non-uniform distributions ($p<0.05$, FDR-corrected) are marked in bold. (b) Average respiratory amplitudes aligned to movement onset across conditions, with inspiration and expiration indicated by falling and rising slopes, respectively. (c) Mean-corrected change in HR around movement onset and (d) HRV power spectral density in the VLF, LF, and HF bands across conditions.

RQ2: We analyzed HR and HRV measures across conditions. HR was elevated during self-paced trials and comparable between cued and resting conditions (Fig. 2c), whereas HRV and LF/HF power were highest during cued trials (Fig. 2c,d), with slightly attenuated peaks for movements cued to expiration compared to inspiration.

RQ3: We examined low-frequency EEG activity time-locked to movement onset. Grand-average signals over frontocentral channels showed a typical MRCP pattern, including a gradual readiness potential, a main negative motor potential around movement onset, and subsequent positive peaks (Fig. 3). Condition effects emerged only after stratifying participants by mean self-paced inter-movement interval (> 6 s vs. < 6 s; n = 10 per group). Participants with longer intervals showed higher MRCP amplitudes during self-paced inspiration and lower amplitudes during cued expiration, whereas those with shorter intervals showed the opposite pattern, with larger amplitudes during cued expiration. Moreover, MRCP peak amplitude correlated positively with HR only for movements cued to expiration ($p<0.05$, FDR-corrected).

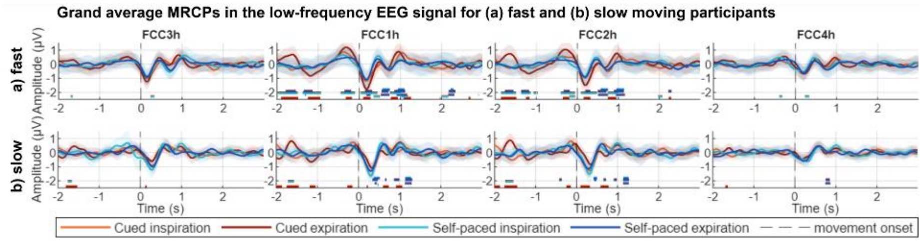


Fig. 3: Grand-average MRCPs aligned to movement onset (t = 0 s) for both cued and self-paced conditions, shown for participants with (a) slow and (b) fast self-paced movements. Signals are displayed over left- and right-hemispheric frontocentral channels (FCC3h, FCC1h – left; FCC2h, FCC4h – right). Horizontal bars indicate time intervals with significant differences between conditions ($p<0.05$, FDR-corrected), with colors corresponding to the compared conditions.

Discussion and Limitations

By simultaneously examining EEG, respiration, and ECG during self-paced and breathing-cued movements, we demonstrate that movement intention is shaped by cardiorespiratory state rather than arising as an isolated cortical event.

First, we observed that movements preferentially occurred during early to mid-expiration (RQ1; Fig. 2a), consistent with previous findings [18]. This pattern likely reflects reduced competition with inspiratory motor control, as actions were avoided near the expiration–inspiration transition, a phase that may require increased neural resources [18,33].

Cardiac responses around movement onset (RQ2) revealed integrated effects of breathing and task demands. HR increased only during self-paced movements, whereas cued conditions maintained near-rest HR levels despite higher attentional load, likely due to constraints imposed by paced breathing. This interpretation is supported by increased LF and HF power during cued trials. Although traditionally linked to sympathetic and vagal activity, respectively [34], their concurrent increase has been associated with cognitive effort and fatigue, and may here reflect heightened attention rather than reduced stress [35,36,37]. HRV around movement onset was dominated by RSA, with movement-related modulations either amplifying or attenuating its effects depending on whether movements were cued to inspiration or expiration.

Regarding cortical responses (RQ3), attention-demanding conditions, particularly cued expiration, produced enhanced MRCP amplitudes and stronger correlations between MRCP amplitude and HR. These effects were most prominent among participants with precise cue adherence, likely driven by active expiratory control and increased cognitive effort. In contrast, slow self-paced movements better reflected intrinsic physiological interactions with task-induced modulation. This suggests that analyzing physiology-driven effects requires experimental conditions that closely match the natural pacing of a participant. Otherwise, the observations may be overshadowed by stronger paradigm-driven effects, such as stress-related responses. Under cued conditions, one possible approach to increase user compliance and reduce cognitive load is real-time adaptation of cue timing to the participant's preferred pace.

It is important to note that several limitations should be considered when interpreting these findings.

A key limitation concerns the translation of the present MRCP-related results into real-world applications. The observed effects were derived from low-amplitude, slow cortical potentials, which require precise time-locking and averaging across trials to achieve a sufficient signal-to-noise ratio. This limits their detectability in single-trial EEG under typical conditions. At the same time, this does not imply that both movement preparation and ongoing movement-related processes cannot be inferred from continuous EEG. While classical MRCP analysis indeed relies on averaging, a substantial body of BCI research has shown that movement intention/action can also be decoded at the single-trial level using appropriate feature extraction and machine learning approaches, including slow cortical potentials, low-frequency time-domain features, and oscillatory signatures such as ERD/ERS [38,39,40]. However, such approaches are inherently more challenging and typically less robust than ERP-based analyses. In this context, our results suggest that the coupling between neural activity and more readily observable cardiorespiratory signals could support MRCP detection or provide complementary information in continuous BCI settings. Importantly, the present study focused on time-domain analyses. Respiration-related effects may also manifest in the frequency domain, including ERD/ERS patterns, which may be more suitable for online detection; however, this remains to be investigated.

More generally, it is also unclear to what extent the observed effects hold across individuals and experimental conditions. This concerns the role of the airflow pathway, voluntary vs. artificial ventilation, and inter-individual differences in physiological fitness and respiratory capacity. Future studies with more homogeneous samples are needed to investigate these factors more systematically.

Finally, although this study examined how respiration influences neural and cardiac dynamics, the direction of these relationships is unclear. Breathing may be influenced by action preparation, or all systems may be jointly modulated by higher-level control mechanisms. Clarifying these bidirectional interactions will be an important direction for future research.

Outlook

In summary, voluntary movement emerges in close interaction with ongoing physiological processes. Breathing phase and HR provide complementary information beyond the readiness potential: natural conditions reveal intrinsic physiological interactions but produce weaker neural responses, whereas task-focused states, such as controlled active expiration, enhance physiological responses and improve detection of intended actions.

In the context of BCI applications, these insights are particularly relevant for asynchronous paradigms involving self-initiated movements, where reliable intention detection is challenging. In such settings, cardiorespiratory signals may provide additional context for detection or be used to modulate neural responses, for instance, through controlled breathing patterns, to improve robustness. While not intended as a standalone solution, these findings can inform the design of future systems that integrate physiological context. Subsequent research should investigate whether incorporating such information, or actively guiding breathing behavior, leads to measurable improvements in real-time detection performance.

These findings may also benefit neurorehabilitation settings, where enhancing the detectability and consistency of movement-related brain signals is essential. Guided breathing could be explored as a simple and non-invasive strategy to modulate neural activity and support training or recovery processes.

In the broader context of neuroinformation systems, cardiorespiratory signals offer a window into the user's internal cognitive and physiological state, enabling more adaptive and context-aware system behavior. Importantly, this perspective goes beyond passive sensing, as physiological processes may also be actively shaped, for example, through guided breathing, to influence neural responses and interaction outcomes. This points toward interactive, human-in-the-loop systems in which both the user and the system co-adapt over time. Within such systems, personalization becomes crucial to ensure that adaptive mechanisms reflect the user's intrinsic physiological dynamics rather than imposing a fixed interaction protocol.

References

1. B. J. Edelman *et al.*, "Non-Invasive Brain-Computer Interfaces: State of the Art and Trends," *IEEE Rev. Biomed. Eng.*, vol. 18, pp. 26–49, 2025, doi: 10.1109/RBME.2024.3449790.
2. H. H. Kornhuber and L. Deecke, "Hirnpotentialänderungen beim Menschen vor und nach Willkürbewegungen, dargestellt mit Magnetband-Speicherung und Rückwärtsanalyse," *Pflugers Arch*, vol. 281, Jan. 1964.
3. H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?," *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2341–2356, Nov. 2006, doi: 10.1016/j.clinph.2006.04.025.
4. T. Parviainen, P. Lyyra, and M. S. Nokia, "Cardiorespiratory rhythms, brain oscillatory activity and cognition: review of evidence and proposal for significance," *Neuroscience & Biobehavioral Reviews*, vol. 142, p. 104908, Nov. 2022, doi: 10.1016/j.neubiorev.2022.104908.
5. T. Engelen, M. Solcà, and C. Tallon-Baudry, "Interoceptive rhythms in the brain," *Nat Neurosci*, vol. 26, no. 10, pp. 1670–1684, Oct. 2023, doi: 10.1038/s41593-023-01425-1.
6. J. L. Herrero, S. Khuvis, E. Yeagle, M. Cerf, and A. D. Mehta, "Breathing above the brain stem: volitional control and attentional modulation in humans," *Journal of Neurophysiology*, vol. 119, no. 1, pp. 145–159, Jan. 2018, doi: 10.1152/jn.00551.2017.

7. D. S. Kluger and J. Gross, “Depth and phase of respiration modulate cortico-muscular communication,” *NeuroImage*, vol. 222, p. 117272, Nov. 2020, doi: 10.1016/j.neuroimage.2020.117272.
8. M. C. Melnychuk, I. H. Robertson, E. R. G. Plini, and P. M. Dockree, “A Bridge Between the Breath and the Brain: Synchronization of Respiration, a Pupillometric Marker of the Locus Coeruleus, and an EEG Marker of Attentional Control State,” Sep. 06, 2021. doi: 10.20944/preprints202109.0092.v1.
9. D. S. Kluger and J. Gross, “Respiration modulates oscillatory neural network activity at rest,” *PLoS Biol*, vol. 19, no. 11, p. e3001457, Nov. 2021, doi: 10.1371/journal.pbio.3001457.
10. O. Perl *et al.*, “Human non-olfactory cognition phase-locked with inhalation,” *Nat Hum Behav*, vol. 3, no. 5, pp. 501–512, Mar. 2019, doi: 10.1038/s41562-019-0556-z.
11. M. Johannknecht and C. Kayser, “The influence of the respiratory cycle on reaction times in sensory-cognitive paradigms,” *Sci Rep*, vol. 12, no. 1, p. 2586, Feb. 2022, doi: 10.1038/s41598-022-06364-8.
12. C. Harting, L. Hehemann, L. Stetza, and C. Kayser, “Respiration shapes response speed and accuracy with a systematic time lag,” *Proc. R. Soc. B.*, vol. 292, no. 2044, p. 20242566, Apr. 2025, doi: 10.1098/rspb.2024.2566.
13. S. Li and W. Z. Rymer, “Voluntary Breathing Influences Corticospinal Excitability of Nonrespiratory Finger Muscles,” *Journal of Neurophysiology*, vol. 105, no. 2, pp. 512–521, Feb. 2011, doi: 10.1152/jn.00946.2010.
14. B. Raßler, U. Bradl, and H.-C. Scholle, “Interactions of breathing with the postural regulation of the fingers,” *Clinical Neurophysiology*, vol. 111, no. 12, pp. 2180–2187, Dec. 2000, doi: 10.1016/S1388-2457(00)00483-1.
15. B. Raßler, “Mutual nervous influences between breathing and precision finger movements,” *European Journal of Applied Physiology*, vol. 81, no. 6, pp. 479–485, Mar. 2000, doi: 10.1007/s004210050071.
16. M. Grund *et al.*, “Respiration, Heartbeat, and Conscious Tactile Perception,” *J. Neurosci.*, vol. 42, no. 4, pp. 643–656, Jan. 2022, doi: 10.1523/JNEUROSCI.0592-21.2021.
17. B. Raßler, I. Nietzold, and S. Waurick, “Phase-dependence of breathing and finger tracking movements during normocapnia and hypercapnia,” *Eur J Appl Physiol*, vol. 80, no. 4, pp. 324–332, Aug. 1999, doi: 10.1007/s004210050599.
18. H.-D. Park, C. Barnoud, H. Trang, O. A. Kannape, K. Schaller, and O. Blanke, “Breathing is coupled with voluntary action and the cortical readiness potential,” *Nat Commun*, vol. 11, no. 1, p. 289, Feb. 2020, doi: 10.1038/s41467-019-13967-9.
19. A. Schurger, M. Mylopoulos, and D. Rosenthal, “Neural Antecedents of Spontaneous Voluntary Movement: A New Perspective,” *Trends in Cognitive Sciences*, vol. 20, no. 2, pp. 77–79, Feb. 2016, doi: 10.1016/j.tics.2015.11.003.
20. G. G. Berntson, J. T. Cacioppo, and K. S. Quigley, “Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications,” *Psychophysiology*, vol. 30, no. 2, pp. 183–196, Mar. 1993, doi: 10.1111/j.1469-8986.1993.tb01731.x.
21. S.-M. Hsu, C.-H. Tseng, C.-H. Hsieh, and C.-W. Hsieh, “Slow-paced inspiration regularizes alpha phase dynamics in the human brain,” *Journal of Neurophysiology*, vol. 123, no. 1, pp. 289–299, Jan. 2020, doi: 10.1152/jn.00624.2019.

22. G. Florian and G. Pfurtscheller, "Elimination von Atmungseffekten auf bewegungsinduzierte Änderungen der Herzrate - Elimination of Respiratory Effects on Movement-induced Cardiac Response," *Biomedizinische Technik/Biomedical Engineering*, vol. 42, no. 7–8, pp. 203–206, 1997, doi: 10.1515/bmte.1997.42.7-8.203.
23. R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, Mar. 1971, doi: 10.1016/0028-3932(71)90067-4.
24. M. Lajtos, L. A. Barradas-Chacón, and S. C. Wriessnegger, "Effects of handedness on brain oscillatory activity during imagery and execution of upper limb movements," *Front. Psychol.*, vol. 14, Jun. 2023, doi: 10.3389/fpsyg.2023.1161613.
25. C. Kothe *et al.*, "The lab streaming layer for synchronized multimodal recording," *Imaging Neuroscience*, vol. 3, p. IMAG.a.136, Sep. 2025, doi: 10.1162/IMAG.a.136.
26. J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
27. The MathWorks, Inc., "pwelch — Welch's power spectral density estimate." Accessed: Jan. 15, 2026. [Online]. Available: https://www.mathworks.com/help/signal/ref/pwelch.html
28. P. D. Welch, "The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967, doi: 10.1109/TAU.1967.1161901.
29. A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/j.jneumeth.2003.10.009.
30. Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, pp. 289–300, Jan. 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.
31. P. Berens, "**CircStat**: A *MATLAB* Toolbox for Circular Statistics," *J. Stat. Soft.*, vol. 31, no. 10, 2009, doi: 10.18637/jss.v031.i10.
32. T. M. Inc, *MATLAB version: 25.1.0.2973910 (R2025a)*. (2025). The MathWorks Inc., Natick, Massachusetts, United States. [Online]. Available: https://www.mathworks.com
33. L. Oostwoud Wijdenes, R. B. Ivry, and P. M. Bays, "Competition between movement plans increases motor variability: evidence of a shared resource for movement planning," *Journal of Neurophysiology*, vol. 116, no. 3, pp. 1295–1303, Sep. 2016, doi: 10.1152/jn.00113.2016.
34. F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Health*, vol. 5, p. 258, Sep. 2017, doi: 10.3389/fpubh.2017.00258.
35. C. Finucane, G. Boyle, C. W. Fan, D. Hade, L. Byrne, and R. A. Kenny, "Mayer wave activity in vasodepressor carotid sinus hypersensitivity," *Europace*, vol. 12, no. 2, pp. 247–253, Feb. 2010, doi: 10.1093/europace/eup385.
36. C. Julien, "The enigma of Mayer waves: Facts and models," *Cardiovascular Research*, vol. 70, no. 1, pp. 12–21, Apr. 2006, doi: 10.1016/j.cardiores.2005.11.008.

37. B. J. Janssen, S. C. Malpas, S. L. Burke, and G. A. Head, “Frequency-dependent modulation of renal blood flow by renal nerve activity in conscious rabbits,” *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 273, no. 2, pp. R597–R608, Aug. 1997, doi: 10.1152/ajpregu.1997.273.2.R597.
38. P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, “Upper limb movements can be decoded from the time-domain of low-frequency EEG,” *PLoS ONE*, vol. 12, no. 8, p. e0182578, Aug. 2017, doi: 10.1371/journal.pone.0182578.
39. M. R. Crell, K. Kostoglou, K. Sterk, and G. R. Müller-Putz, “A novel paradigm for fast training data generation in asynchronous movement-based BCIs,” *Front. Hum. Neurosci.*, vol. 19, p. 1540155, Feb. 2025, doi: 10.3389/fnhum.2025.1540155.
40. M. R. Crell, K. Kostoglou, P. Suwandjieff, J. R. D. Cruz, and G. R. Müller-Putz, “A Non-Invasive, MRCP-Based BCI for Online Communication,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 34, pp. 1272–1283, 2026, doi: 10.1109/TNSRE.2026.3666564.

Brain Speaks Louder Than Words: Neurophysiological Opinion Inference via Cognitive Dissonance Triggers

Qi Zhao[1], Elena Karahanna[2], D.J. Wu[3], and Min Ding[4]

[1] University of New South Wales, Sydney, Australia
qi.zhao4@unsw.edu.au
[2] University of Georgia, Athens, USA
ekarah@uga.edu
[3] Georgia Institute of Technology, Atlanta, USA
dj.wu@scheller.gatech.edu
[4] Penn State University, University Park, USA
minding@psu.edu

Abstract. Contemporary digital businesses and services demand real-time, authentic understanding of user opinions, yet traditional approaches such as surveys and behavior-based inference suffer from response biases, data sparsity, and privacy concerns. To address these challenges, we design, instantiate, demonstrate, and evaluate a novel NeuroIS artifact in this research. Building on the heuristic-systematic model and cognitive dissonance theory, we propose an innovative artifact that infers individual users' authentic opinions directly from neural indicators when users are presented with artificial stimuli incongruent with their opinions. This incongruence triggers cognitive dissonance and encourages more effortful systematic information processing, which provides reliable neural indicators of authentic opinion. Our empirical validation demonstrates significant neural indicators of increased systematic processing in response to incongruent stimuli exposure, with this effect moderated by a heuristic cue. Relative to benchmark methods, our artifact shows an improved out-of-sample predictive performance in predicting individuals' opinions.

Keywords: Opinion inference · Cognitive dissonance · Information processing · Design science · NeuroIS · EEG

Introduction

Gaining reliable, real-time insights into individual user opinions has become a central challenge for designing, delivering, and governing digital businesses and services. Foundational Information Systems (IS) domains, from user experience and personalization to platform governance and risk communication, depend on accurate and timely assessments of user opinion and preference [9, 17, 28, 43, 44, 47]. However, existing tools predominantly used for this purpose face growing limitations for IS practitioners and researchers. Traditional self-report instruments, such as surveys, are subject to well-established response biases, disrupt continuous user experiences, and introduce latency between data collection and application [15, 29, 48]. Meanwhile, inferring opinion from behavioral trace data (e.g., clickstream, browsing history, purchase records) is constrained by inference bias, data sparsity, and an increasingly complex privacy landscape [1, 26]. In this research, we follow the design science paradigm [18, 22, 23, 36] to develop a theory-driven neuroscientific artifact that addresses the following question: *How can we infer authentic individual opinions unobtrusively and in near-real time while protecting user privacy and requiring minimal calibration?*

Specifically, we propose a novel NeuroIS artifact: Electroencephalography Dissonance-Driven Opinion Inference Artifact (EDOIA). Rather than attempting to directly decode the neural representation of opinions as in existing studies [21, 42], our approach builds on a strong theoretical foundation: we infer opinions by detecting the neurophysiological footprint of artificially induced cognitive dissonance. Drawing upon the heuristic-systematic model (HSM, [5, 6, 34], and cognitive dissonance theory (CDT) as our kernel theories [13, 33], we posit that presenting users with stimuli incongruent with their authentic opinions induces dissonance (i.e., an incongruent claim about the user's favored service or product), which in turn triggers effortful systematic processing [5, 25, 31, 33]. The neurophysiological correlate of this enhanced cognitive effort is a well-documented desynchronization of EEG activity in the alpha and beta frequency bands [20, 32, 37], referred to as event-related desynchronization (ERD, [27]). EDOIA instantiates and operationalizes this mechanism through a sequential stimulus presentation protocol. Specifically, immediately after exposing a user to the focal evaluating object stimulus (e.g., a product image or news headline for which the user's opinion is of interest), EDOIA presents *artificial stimuli* expressing extreme opinions that may contradict the user's authentic opinion of the object stimulus. By measuring time-locked ERD immediately following the artificial stimulus presentation, EDOIA infers whether the valence of the user's underlying opinion is opposite to the artificial stimulus, thus eliciting greater systematic processing as suggested by the kernel theories.

We demonstrate and evaluate the proposed EDOIA through an empirical study. Our empirical validation demonstrates that exposure to incongruent stimuli induces robust ERD in EEG alpha/beta band power, indicative of more systematic processing. When a strong heuristic cue is easily accessible, the following moderation analysis reveals a significant attenuation of ERD. Guided by the observed ERD patterns, we show that EDOIA provides a significant improvement in out-of-sample individual opinion prediction compared with the benchmark methods. The prediction comparison also delineates the boundary conditions of EDOIA's performance, consistent with the moderation effect: EDOIA excels in prediction only when the heuristic cue is less salient.

Design of EDOIA

Kernel Theories and Neuroscientific Foundations

The design of EDOIA draws on dual-process theories of information processing, specifically HSM, which operates primarily during the retrieval and judgment stages [6, 7, 34]. The model posits two distinct cognitive routes, heuristic and systematic, through which individuals engage cognitively with presented information. The heuristic route involves quick, automatic, and often subconscious mental shortcuts, allowing individuals to make rapid judgments or decisions based on the availability of simplified cues [30]. In contrast, the systematic route involves more deliberate and comprehensive retrieval and processing of information, entailing careful evaluation, critical thinking, and deeper scrutiny of details and evidence [6, 7, 16].

Cognitive dissonance theory (CDT) provides the second theoretical pillar for our artifact design. According to CDT, exposure to information incongruent with an individual's cognitions, beliefs, attitudes, or behaviors causes psychological tension [13, 35, 41]. The experienced dissonance lowers individuals' confidence in their judgments and motivates them to mitigate dissonance through various means [30, 41]. One major route involves acquiring more specific information about the focal object and conducting more systematic processing to resolve the incongruency, a route represented by systematic processing [6, 31].

In EEG studies, raw EEG signals are typically decomposed into electrical power level (dB) across five frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-45 Hz). These power levels across different bands have been associated with different behavioral or cognitive tasks. Especially, ERD of alpha/beta band power around the frontal cortex, measured as a substantial reduction in power level, has been linked to enhanced cognitive processing across cognitive tasks and even species [20, 33, 37, 38, 39]. For example, reduction of alpha/beta band power can serve as a proxy for more stimulus-specific information processing [20]. Similarly, when users are asked to rate articles using self-referential questions, desynchronization of alpha/band power in the frontal cortex indicates enhanced systematic processing [33].

These kernel theories and neuroscientific foundations suggest the following mechanism: exposure to incongruent information can induce cognitive dissonance and motivate users to commit greater cognitive effort to retrieving and processing information related to the focal object. This enhanced cognitive processing can cause desynchronization and a substantial reduction of EEG alpha/beta band power in the frontal cortex.

Sequential Presentation Protocol Design

Accordingly, we devise and include a set of *artificial stimuli* with known valence to induce cognitive dissonance through incongruency. By design, these artificial stimuli express strongly positive or negative opinions that can be either congruent or incongruent with users' authentic opinions toward the object stimuli. When the presented artificial stimulus is incongruent with the individual's opinion, we expect it to challenge the individual's judgmental confidence and induce more systematic processing. An inference of the user's authentic opinion can then be made as that with the opposite valence of the presented artificial stimulus.

With two sets of stimuli involved, EDOIA relies on how users cognitively process the artificial stimuli to generate inferences. To cleanly isolate the effect of artificial stimuli and control variation both within and across stimulus exposures, we devise a sequential presentation procedure that first presents the object stimuli, followed by artificial stimuli. Specifically, one exposure trial starts with presenting only the object stimulus, followed by the presentation of both the object stimulus and the artificial stimulus together. Figure 1 is an illustrative example of the proposed presentation protocol in our lab study to study individual food preferences (Chinese cuisine dishes).

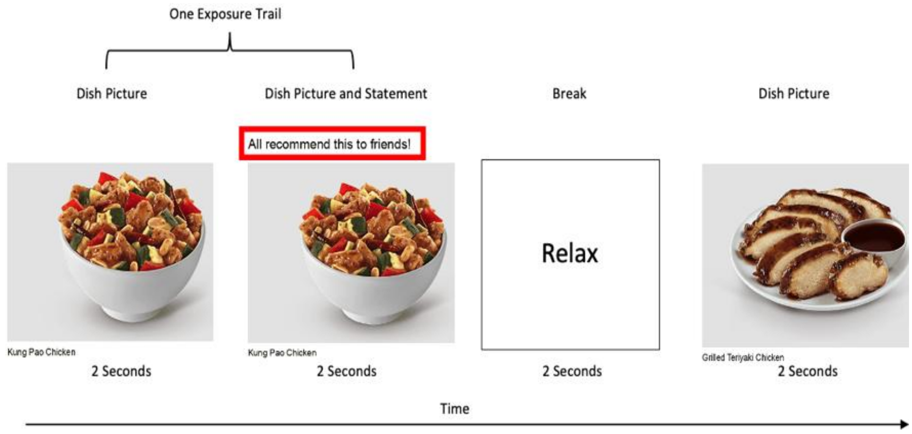


Fig. 1. The sequential stimuli presentation protocol

Event Study Adaption

With the proposed sequential stimulus protocol, measurement of power change in the alpha/beta band upon artificial stimulus exposure requires careful analytical attention. In this study, we adapt the event study methodology [24, 40]. Adaptation of the event study methodology in our context effectively controls for within-exposure-trial variation and provides a robust measure of the artificial stimulus exposure effect.

Figure 2 illustrates the event study adaptation for only one exposure trial. For each focal user i and object d pair, we first fit a multivariate linear model of actual band powers ($\mathbf{E}_{\mathrm{idet}}$) using the baseline activity ($\bar{\mathbf{E}}_{\mathrm{dt},-\mathrm{i}}$) as predictor, derived by averaging other users' band powers viewing the object d, over the object stimulus only period. Then, we calculate the difference between the actual band powers ($\mathbf{E}_{\mathrm{idet}}$) and the expected powers ($\hat{\mathbf{E}}_{\mathrm{idet}}$) derived from the baseline activity ($\bar{\mathbf{E}}_{\mathrm{dt},-\mathrm{i}}$) to measure the power change in the alpha/beta band after artificial stimuli exposure. Following the event study literature, we refer to this difference at each time step as *abnormal activity* (AA). Finally, we cumulate the derived AAs over the object and artificial stimuli period to obtain the *cumulative abnormal activity* (CAA) as an aggregated measure of the reduction of EEG alpha/beta band power induced by artificial stimuli exposure.

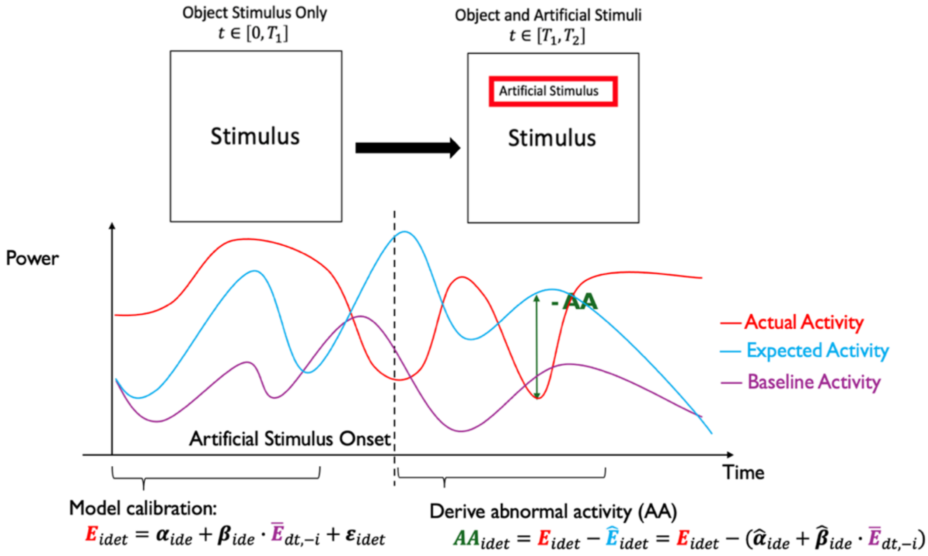


Fig. 2. The devised sequential stimuli presentation protocol
Notes: The three EEG signal lines in this figure are for illustration only and are not real data. The specific AA derived in the figure is negative (-), namely, a reduction in alpha/beta band power, which indicates a more systematic processing.

Opinion Inference Rules by CAAs

Once the artificial stimulus induced CAAs are obtained, we use them as focal ERD indicators and infer an individual user's opinion based on the nature of the presented incongruency. Specifically, a negative CAA (i.e., a reduction in alpha/beta band power) following artificial stimulus exposure suggests more systematic processing of stimulus-specific information, indicating that the artificial stimulus is incongruent with the individual user's authentic opinion toward the focal object. We thus infer the valence of the authentic opinion to be opposite to that of the presented artificial stimulus. When both positive and negative artificial stimuli produce negative CAAs, the final inference is based on the stronger dissonance reaction, the negative CAA with greater magnitude.

Empirical Demonstration and Evaluation

Following the design science paradigm [18, 22, 23, 36], we implement a lab study to demonstrate and evaluate the proposed EDOIA. The objectives are to validate EDOIA's core underlying mechanism and demonstrate that leveraging artificially induced ERD patterns can improve individual opinion inference from existing methods. This section describes the study design and validation evidence. Specifically, we focus on inferring individual users' preferences over a set of products, an important context for user opinions for IS practitioners and scholars.

Study Design

Eighty-four participants were recruited from the business undergraduate population at an eastern US public university in exchange for extra credit toward enrolled courses. Prior screening ensured participants had no history of neurological illness or damage, were not using any psychiatric medication, and had normal or corrected-to-normal vision. The study proceeded after obtaining informed consent from participants.

The object stimuli set includes product items toward which participants' authentic preferences are of interest. In this study, we selected Chinese food as our focal product category. First, the college town hosting the university features approximately a dozen Chinese restaurants. Specifically, the university student union center, located centrally on campus, houses a Panda Express, which is among the most popular on-campus dining options for undergraduate students. Second, because Chinese dishes tend to feature more diverse and exotic flavors and ingredients, they naturally induce more heterogeneous preferences among American students. The final object stimuli set thus includes 12 regular dishes served at Panda Express and 18 of the most popular dishes served at another popular Chinese restaurant adjacent to campus. We collected images of these 30 dishes from the restaurants' websites as our object stimuli.

Although our proposed framework allows generalizable implementation, we adopt a simple design and devise the artificial stimuli in this study as textual statements expressing strong opinions. Because individuals' authentic preferences toward these dishes, either like or dislike, are unknown a priori, we devise statement stimuli with both positive (e.g., "People always LIKE this!") and negative opinions (e.g., "No one LIKES this!"), such that when participants truly like (dislike) a dish, exposure to a strong negative (positive) statement is expected to elicit dissonance through incongruency.

EEG Data Collection and Preprocessing

EEG data were collected using the Emotiv EPOC X headset[1]. The headset is a 14-channel system used in prior NeuroIS studies published in leading journals [32, 33, 34]. Compared with traditional non-invasive EEG equipment used for medical diagnostic purposes, this device is significantly more portable and easier to mount on participants' scalps [10]. Neuroscience and NeuroIS researchers have widely scrutinized the Emotiv device and found that this consumer-grade system provides high temporal resolution EEG data as accurate and reliable as medical-grade systems [33].

Among 14 EEG electrode locations dispersed over the scalp along the classic 10-20 system, we use all six locations in the prefrontal cortex area (AF3, AF4, F3, F4, F7, F8). Numerous neuroscientific studies have demonstrated a strong correlation between neural activity of the prefrontal/frontal cortex and implicit valuation, preference processing, actual choices, and population-wide commercial performance [28, 33, 34, 42]. Additionally, the locations of prefrontal/frontal EEG sensors align closely with the contact points between most digital wearable headsets and human skin. This means that these devices can easily incorporate EEG sensors to implement EDOIA for near-future commercialization.

[1] https://www.emotiv.com/epoc-x

Following the literature, the raw EEG data were first down-sampled to 128 Hz and filtered with a passband from 1 Hz to 49 Hz. The EEG data were then average-referenced, and epochs were extracted for a window of 500 ms pre-stimuli to 4000 ms after the object stimulus onset, with the 500 ms to 100 ms pre-stimuli period used for baseline correction. To remove artifacts related to muscle and eye movements, independent component analysis was conducted, and epochs impacted significantly by muscle activities were removed from further analysis. We then derived the ERSP time-frequency presentation of EEG data with five bands.

Empirical Results

We first classify all exposure trials in which participants like (dislike) the presented dish, followed by negative (positive) artificial statements, as "Incongruent" exposures, while exposure trials in which the presented statements match participant preferences are classified as "Congruent" exposures. The true labels of liking/disliking are based on participants' self-reported preference scales.

Table 1 presents the results of both parametric and nonparametric tests for the two groups of CAAs, congruent or incongruent. Consistent with the theoretical foundations, we find a significant reduction in alpha/beta band power only when the presented statement is incongruent with the participant's preference, confirming that artificially induced dissonance encourages individuals to engage in more systematic processing of the presented stimuli.

Table 1. Test statistics of statement exposure – main effect.

	Tests	Congruent (p-value)	Incongruent (p-value)
Parametric	Cross-sectional Test (t)	0.02 (0.98)	**-3.21 (0.00)** ***
	Standardized Cross-sectional Test (t)	0.05 (0.96)	**-3.49 (0.00)** ***
Nonparametric	Generalized Sign Test (z)	0.18 (0.86)	**-4.11 (0.00)** ***
	Generalized Rank Test (t)	-0.06 (0.95)	**-4.45 (0.00)** ***

Responding to calls in design science research to articulate design knowledge rather than artifact novelty alone [18], we further explore the potential boundary conditions. Specifically, we examine the moderating effect of an accessible heuristic cue, whether there is an opinion consensus across the population on each dish. Table 2 presents the results for the 2 × 2 combinations (Congruency × Consensus). Results demonstrate that the dissonance-induced reduction of alpha/beta band power is statistically significant when no strong consensus exists on the focal dish. In other words, an extreme statement induces a more substantial dissonance effect for dishes on which more divergent opinions exist among individuals, such that exposure to extreme statements becomes less expected. Consistent with previous studies, we find that individuals tend to adopt a less systematic approach to processing information when a consensus cue is available, as indicated by a less pronounced reduction in alpha/beta band power.

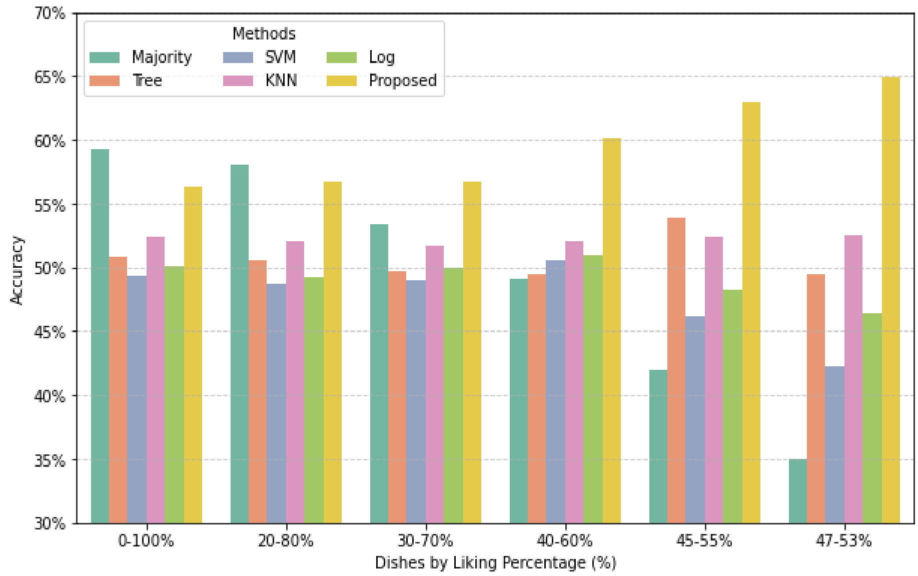
Table 2: Test statistics of statement exposure – moderating effect of heuristic cue.

	Tests	Consensus	Congruent (p-value)	Incongruent (p-value)
Parametric	Cross-sectional Test (t)	Yes	0.43 (0.67)	-0.73 (0.47)
		No	-0.36 (0.72)	**-3.63 (0.00)** ***
	Standardized Cross-sectional Test (t)	Yes	0.58 (0.56)	-1.09 (0.28)
		No	-0.43 (0.67)	**-3.65 (0.00)** ***
Nonparametric	Generalized Sign Test (z)	Yes	0.03 (0.98)	-1.29 (0.20)
		No	0.21 (0.83)	**-4.36 (0.00)** ***
	Generalized Rank Test (t)	Yes	0.19 (0.85)	-1.08 (0.28)
		No	-0.27 (0.79)	**-4.11 (0.00)** ***

Finally, we evaluate whether EDOIA can provide reliable preference inferences at the individual level, by comparing its out-of-sample predictive performance with five benchmark methods: (1) majority rule ("Majority"), using the majority opinion from other users to infer the focal user's opinion, which simulates the common practice of recommending "best-seller" products or services. The second set of benchmarks represents existing EEG-based preference inference studies and includes four classical machine learning methods [21]: (2) support vector machine (SVM) with a linear kernel; (3) logistic regression (Log); (4) AdaBoost decision trees (Tree), with 300 trees and a minimum leaf size of 5 for regularization; and (5) K-nearest neighbors (KNN) with K = 5.

Without the proposed stimulus presentation protocol in EDOIA, current EEG studies derive various features from raw EEG data collected during the object-stimulus-only period in Figure 2. For the benchmark methods, we assume no access to EEG data after artificial stimulus exposure and follow their practices to derive the following 15 EEG features as the predictors: five frequency band powers over the frontal area, hemispheric asymmetry features across five frequency bands, and inter-individual correlations across five frequency bands (refer to [21] for full details).

Figure 3 presents the prediction results across six groups of dishes with gradually decreasing consensus (the number of dishes "# Dish", and the number of observations "# Obs", within each group are provided at the bottom of the figure). As Figure 3 shows, EDOIA performs significantly better than other benchmark machine learning methods using EEG data. Although the difference between EDOIA and the majority benchmark is not significant when all dishes are considered, the improvement from EDOIA over the majority rule becomes substantial and significant for dishes with a less salient consensus cue (i.e., starting from 40%-60% liking percentage). These results verify that the proposed artifact improves individual preference inference performance relative to existing methods when dissonance-induced systematic processing is salient.



# Dish	30	25	21	12	5	3
# Obs	511	496	406	271	143	97

Fig. 3. Predictive performance across competing methods

Discussion and Concluding Remarks

This research contributes to NeuroIS and the broader IS literature by developing and evaluating the EDOIA, a theory-grounded artifact for unobtrusive, individual-level opinion inference [10, 11, 12, 14, 19, 34, 35, 38, 39]. Our study serves as a proof-of-concept design science effort, presenting the findings that establish initial feasibility and theoretical plausibility. Framed through the lens of design science research, our contribution lies in the purposeful construction and initial evaluation of an artifact intended to address a relevant IS problem while generating design knowledge that can inform neuro-adaptive systems in future [18, 22, 36, 46].

First, this study contributes a novel stimulus protocol that translates kernel theories of CDT and HSM into an actionable NeuroIS design. By sequentially presenting a focal object followed by artificial stimuli, the protocol creates theoretically motivated incongruence that is associated with measurable ERD in the alpha/beta bands. The contribution extends beyond a single empirical implementation and offers a reusable design principle for eliciting neurophysiological signals that are informative for opinion inference in digitally mediated settings.

Second, we instantiate this protocol in a deployable artifact, an unobtrusive system that avoids self-report biases, preserves user privacy, and incorporates a key methodological innovation for individual-level inference. The proposed event study adaptation captures cognitive dynamics as they occur and eliminates the need for extensive, person-specific calibration samples required by traditional BCI designs.

Third, our evaluation yields initial knowledge about the boundary conditions under which the artifact is likely to be effective. The findings indicate that opinion-incongruent stimuli are associated with stronger ERD and that this relationship is attenuated when strong heuristic cues such as consensus are present. These results are consistent with the theorized mechanism and provide support for the explanatory value of the underlying kernel theories, while also clarifying when neurophysiological signals may improve individual-level inference. In this respect, the paper contributes not only an artifact, but also design principles and boundary conditions [18, 36].

Digital service and user engagement now unfold in seconds, yet IS practitioners and researchers mostly rely on self-reports and behavioral logs to infer what individual users value. Our EEG-based opinion inference may prove valuable in near real time settings where behavioral records are limited, and where self-reports are biased, strategically distorted, or impractical, in both academia [2, 4, 8] and industry [3, 45]. However, we acknowledge that the proposed artifact should be evaluated across more diverse populations, decision contexts, and hardware configurations, and under field conditions that more closely reflect real-world digital service environments.

References

1. Aggarwal, C. C.: Recommender Systems. Cham: Springer International Publishing (2016)
2. Agarwal R., Karahanna E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly. 1, 665–94 (2000)
3. Barnett, S. B., Cerf, M.: A Ticket for Your Thoughts: Method for Predicting Content Recall and Sales Using Neural Similarity of Moviegoers. Journal of Consumer Research. 44, 160–181 (2017)
4. Baumgartner, H., Sujan, M., Padgett, D.: Patterns of Affective Reactions to Advertisements: The Integration of Moment-to-Moment Responses into Overall Judgments. Journal of Marketing Research. 34, 219–232 (1997)
5. Chaiken, S., Ledgerwood, A.: A Theory of Heuristic and Systematic InformationProcessing. In P. Van Lange, A. Kruglanski, E. Higgins, Handbook of Theories of Social Psychology, pp. 246–266. SAGE Publications Ltd. (2012)
6. Chen, S., Chaiken, S.: The Heuristic-Systematic Model in Its Broader Context. In Dual-Process Theories in Social Psychology, pp. 73–96. The Guilford Press (1999)
7. Chen, S., Duckworth, K., Chaiken, S.: Motivated Heuristic and Systematic Processing. Psychological Inquiry. 10, 44–49 (1999)
8. Chen Y., Galletta D.F., Lowry P.B., Luo X., Moody G.D., Willison R.: Understanding Inconsistent Employee Compliance with Information Security Policies Through the Lens of the Extended Parallel Process Model. Information Systems Research. 32, 1043–65 (2021)
9. Chen, Y., Rui, H., Whinston, A. B.: Conversation Analytics: Can Machines Read Between the Lines in Real-Time Strategic Conversations? Information Systems Research. 36, 440–455 (2025)
10. Davis, F. D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 13, 319–340 (1989)
11. Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G., Benbasat, I., Gefen, D.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly. 36, 679–702 (2012)
12. Dimoka, A., Pavlou, P. A., Davis, F. D.: Research Commentary — NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. Information Systems Research. 22, 687–702 (2011)
13. Festinger, L.: Cognitive Dissonance. Scientific American. 207, 93–106 (1962)

14. Fischer, T., Davis, F. D., Riedl, R.: 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. Vol. 29, pp. 1–10. Springer International Publishing (2019)
15. Fisher, R. J.: Social Desirability Bias and the Validity of Indirect Questioning. Journal of Consumer Research. 20, 303–315 (1993)
16. Fiske, S. T., Taylor, S. E.: Social Cognition. Mcgraw-Hill Book Company 1991
17. Ghose, A., Ipeirotis, P. G., Li, B.: Modeling Consumer Footprints on Search Engines: An Interplay with Social Media. Management Science. 65, 1363–1385 (2019)
18. Gregor, S., Hevner, A. R.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly. 37, 337–355 (2013)
19. Gregor, S., Lin, A. C. H., Gedeon, T., Riaz, A., Zhu, D.: Neuroscience and a Nomological Network for the Understanding and Assessment of Emotions in Information Systems Research. Journal of Management Information Systems. 30, 13–48 (2014)
20. Griffiths, B. J., Mayhew, S. D., Mullinger, K. J., Jorge, J., Charest, I., Wimber, M., Hanslmayr, S.: Alpha/beta Power Decreases Track the Fidelity of Stimulus-Specific Information. eLife. 8, (2019)
21. Hakim, A., Klorfeld, S., Sela, T., Friedman, D., Shabat-Simon, M., Levy, D. J.: Machines Learn Neuromarketing: Improving Preference Prediction from Self-Reports using Multiple EEG Measures and Machine Learning. International Journal of Research in Marketing. 38, 770–791 (2021)
22. Hevner, A. R.: A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems, 19, 4 (2007)
23. Hevner, A. R., March, S. T., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly. 28, 75–105 (2004)
24. Im, K. S., Dow, K. E., Grover, V.: Research Report: A Reexamination of IT Investment and the Market Value of the Firm — An Event Study Methodology. Information Systems Research. 12, 103–117 (2001)
25. Jain, S. P., Maheswaran, D.: Motivated Reasoning: A Depth-Of-Processing Perspective. Journal of Consumer Research. 26, 358–371 (2000)
26. Karwatzki, S., Dytynko, O., Trenz, M., Veit, D.: Beyond the Personalization – Privacy Paradox: Privacy Valuation, Transparency Features, and Service Personalization. Journal of Management Information Systems. 34(2), 369–400 (2017)
27. Klimesch W, Schimke H, Doppelmayr M, Ripper B, Schwaiger J, Pfurtscheller G.: Event-Related Desynchronization (ERD) and the DM Effect: Does Alpha Desynchronization During Encoding Predict Later Recall Performance? International Journal of Psychophysiology. 24, 47–60 (1996)
28. Kuan, K. K., Zhong, Y., Chau, P. Y.: Informational and Normative Social Influence in Group-Buying: Evidence from Self-reported and EEG data. Journal of Management Information Systems. 30, 151–178 (2014)
29. Kwan, S. S. K., So, M. K. P., Tam, K. Y.: Research Note — Applying the Randomized Response Technique to Elicit Truthful Responses to Sensitive Questions in IS Research: The Case of Software Piracy Behavior. Information Systems Research. 21, 941–959 (2010)
30. Maheswaran, D., Chaiken, S.: Promoting Systematic Processing in Low-Motivation Settings: Effect of Incongruent Information on Processing and Judgment. Journal of Personality and Social Psychology. 61, 13–25 (1991)
31. McGrath, A.: Dealing with Dissonance: A Review of Cognitive Dissonance Reduction. Social and Personality Psychology Compass. 11, (2017)
32. Minas, R. K., Potter, R. F., Dennis, A. R., Bartelt, V., Bae, S.: Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams. Journal of Management Information Systems. 30, 49–82 (2014)

33. Moravec, P. L., Kim, A., Dennis, A. R., Minas, R. K.: Do You Really Know if It's True? How Asking Users to Rate Stories Affects Belief in Fake News on Social Media. Information Systems Research. 33, 887–907 (2022)
34. Moravec, P. L., Minas, R. K., Dennis, A. R.: Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All. Management Information Systems Quarterly. 43, 1343-A13 (2019)
35. Oshikawa, S.: The Theory of Cognitive Dissonance and Experimental Research. Journal of Marketing Research. 5, 429–430 (1968)
36. Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems. 24, 45–77 (2007)
37. Pfurtscheller, G., Neuper, C., Mohl, W.: Event-Related Desynchronization (ERD) During Visual Processing. International Journal of Psychophysiology. 16, 147–153 (1994)
38. Riedl, R., Davis, F. D., Hevner, A. R.: Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. Journal of the Association for Information Systems. 15, 4 (2014)
39. Riedl, R., Fischer, T., Léger, P.-M., Davis, F. D.: A Decade of NeuroIS Research: Progress, Challenges, and Future Directions. ACM SIGMIS Database: The DATABASE for Advances in Information Systems. 51, 13–54 (2020)
40. Sorescu, A., Warren, N. L., Ertekin, L.: Event Study Methodology in the Marketing Literature: An Overview. Journal of the Academy of Marketing Science. 45, 186–207 (2017)
41. Spangenberg, E. R., Sprott, D. E., Grohmann, B., Smith, R. J.: Mass-Communicated Prediction Requests: Practical Application and a Cognitive Dissonance Explanation for Self-Prophecy. Journal of Marketing. 67(3), 47–62 (2003)
42. Telpaz, A., Webb, R., Levy, D. J.: Using EEG to Predict Consumers' Future Choices. Journal of Marketing Research. 52, 511–529 (2015)
43. Vance, A., Anderson, B., Kirwan, C. B., Eargle, D.: Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (EEG). Journal of the Association for Information Systems. 15(10), (2014)
44. Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., Kirwan, C. B.: Tuning Out Security Warnings. MIS Quarterly. 42, 355–380 (2018)
45. Venkatraman, V., Dimoka, A., Pavlou, P. A., Vo, K., Hampton, W., Bollinger, B., Hershfield, H. E., Ishihara, M., Winer, R. S.: Predicting Advertising success beyond Traditional Measures: New Insights from Neurophysiological Methods and Market Response Modeling. Journal of Marketing Research. 52, 436–452 (2015)
46. Vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., Riedl, R.: Advancing a NeuroIS Research Agenda with Four Areas of Societal Contributions. European Journal of Information Systems. 29, 9–24 (2020)
47. Xiao, L., Wu, D. J., Ding, M.: A Smart Ad Display System. Information Systems Research. 35, 1873–1889 (2024)
48. Zheng, J., Yin, G., Tan, Y., Ding, J.: Does Help Help? An Empirical Analysis of Social Desirability Bias in Ratings. Information Systems Research. 35, 1052–1073 (2024)

The Neural Recruiter: A Research Proposal for Evaluating Human-GenAI Dynamics in High-Stakes Decision-Making

Thimo Schulz[1], Leonardo Banh[2], and Gero Strobel[2]

[1] Karlsruhe Institute of Technology KIT, Institute for Information Systems WIN, Karlsruhe, Germany
thimo.schulz@kit.edu
[2] University of Duisburg-Essen, Rhine-Ruhr Institute of Information Systems, Essen, Germany
{leonardo.banh,gero.strobel}@uni-due.de

Abstract. Generative AI (GenAI) is increasingly used as an interactive decision partner in organizational context, yet its influence on high-stakes moral decisions and accountability on a cognitive as well as neural base remains less researched. We therefore propose a NeuroIS experiment in a recruitment “tie-breaker” scenario in which participants select between similarly qualified candidates under ethically salient trade-offs to investigate the unique human-GenAI dynamics. The study compares a baseline without GenAI decision support against two GenAI interaction styles: a confident, high-normative-force advisor that promotes acceptance, and a questioning, reflection-fostering advisor that prompts scrutiny. Combining EEG and eye tracking with behavioral logs and self-reports, we examine how GenAI changes cognitive control and engagement, reliance on system output, and perceived accountability for decision-making. Our work contributes with process-level evidence on when GenAI induces active collaboration versus blind compliance in high-stakes decision support and promotes insights into the neural correlates of human-GenAI decision dynamics.

Keywords: Generative AI · Moral Decision-Making · Recruitment · EEG · Eye-tracking

Introduction

Generative artificial intelligence is rapidly shifting from a background productivity technology toward an interactive decision partner that can provide suggestions, generate rationales, and shape how people justify choices in real time [1–3]. This raises concerns from an Information Systems (IS) and management perspective that go beyond performance to agency, governance, and responsible use [4–7], particularly in high-stakes moral decisions where outcomes are consequential and normatively contested [8, 9]. Recruitment is a prominent example: hiring decisions are consequential and constrained by equal opportunity, non-discrimination, and procedural fairness [10, 11]. This is especially true in tie-breaker situations, in which two candidates appear similarly qualified, and a final decision must still be made. In such cases, GenAI does not merely automate screening; it may actively shape how decision makers interpret equivalently strong applicant profiles, which criteria they foreground, and how they justify selecting one candidate over the other [12–14].

A core challenge is that GenAI support may either blur responsibility and accountability (e.g., by enabling justificatory deference) or foster deeper reflection (e.g., by surfacing uncertainty and counterarguments). Yet, evidence remains limited on which dynamics dominates and how GenAI reshapes the micro-processes of judgment and justification in fairness-sensitive decisions, particularly from a neurophysiological perspective [8, 15]. Thus, we aim to answer the following research question:

RQ: How does GenAI assistance and its interaction style shape (a) cognitive control/engagement during hiring decisions, (b) reliance on the system, and (c) perceived responsibility/accountability for the final choice?

Because much prior work relies on outcomes and self-reports, it may miss fast, partly nonconscious processes central to fairness-sensitive judgment [16, 17]. NeuroIS methods can capture these in situ via physiological indicators of mental effort and control [18, 19]. We address this gap using a NeuroIS experiment in which participants complete recruitment tie-breaker decisions with and without GenAI support. We combine behavioral and self-report measures with EEG time-frequency indicators of cognitive control/engagement and, paired with eye tracking, of attention allocation to test whether GenAI support is associated with (i) altered cognitive control/engagement, (ii) increased reliance, and (iii) shifts in perceived responsibility/accountability [10, 20–23].

Related Work

AI-Assisted Moral Decision-Making in High-Stakes Environment: In critical domains such as defense, healthcare, and autonomous transportation, the integration of AI decision support systems is reshaping how human operators process information and attribute responsibility [9, 14, 24]. High-stakes environments are characterized by ethical complexity and significant consequences for human life [14]. For instance, AI is increasingly used as a decision support tool for augmenting medical diagnosis decision [25–27]. In organizational contexts like recruitment, AI is utilized to enhance human processing power and promised to mitigate human biases [28, 29]. Yet the delegation to and reliance on AI systems introduces potential conflicts of authority that require joint control mechanisms to ensure fairness and legal compliance [10, 30, 31]. Research indicates that in these morally challenging situations, human decisions are frequently guided by the autonomous system's behavior [32, 33]. This reliance is often amplified by anthropomorphism, which influences users' attitudes towards AI recommendations [34–36]. High-stakes moral dilemma tasks like the "Moral Machine" or the "Trolley" experiment reveal that users perceive AI decisions that conflict with their own moral judgments as highly significant and surprising [8, 9, 37]. However, additional research is needed that investigates how GenAI as a more salient and persuasive technology with natural language capabilities impacts moral decision-making.

Human-GenAI collaboration from a Neurophysiological Perspective: GenAI as a new paradigm offers new pathways to investigate implications on an individual level [38, 39], such as regarding creativity [3, 40–42], in education [43, 44], or regarding knowledge work [45–49]. Neurophysiological studies investigate new cognitive dynamics in human-GenAI collaboration and allow a complementary view of unconscious effects of the human user beyond self-reported data [38, 48]. For instance, a comprehensive EEG study by Kosmyna et al. reveals that collaborating with GenAI poses the potential to alter neural connectivity patterns compared to unassisted effort without GenAI use [15]. Furthermore, an fNIRS study by Fabre et al. on moral decision-making with AI advisors shows that being provided with deontological arguments leads to a decrease in right dorsolateral prefrontal cortex activity, suggesting a voluntary down-regulation of affective responses to preserve one's self-image [8].

Hypotheses and Research Model Development

Building on Technology Dominance Theory (TTD), we conceptualize GenAI as an *intelligent decision aid* whose effects operate primarily through reliance: weighting and following the system's output [50–52]. Reliance is especially likely under ambiguity, when multiple criteria are defensible and the 'correct' choice is uncertain [51, 53]. We extend this perspective by integrating Social Influence Theory (SIT), which explains how the framing of advice affects acceptance of decision support [54]. High-certainty, high-normative-force advice increases compliance-like acceptance, whereas questioning prompts foster reflection and reduce passive agreement [53, 55]. Integrating TTD and SIT, we posit that GenAI affects accountability via (1) in-task cognitive control/engagement during advice integration and (2) resulting changes in reliance, which then shape perceived accountability.

NeuroIS emphasizes that IS artifacts shape outcomes via time-resolved cognitive processes that self-reports may miss, motivating neurophysiological measures of in-task control/engagement [56, 57]. In our paradigm, *questioning and reflection-fostering* GenAI style is expected to surface reflection, increasing conflict monitoring and cognitive control demands during the decision episode. A *confident, high-normative-force* GenAI style is expected to reduce perceived ambiguity and promote acceptance, lowering control demands and encouraging acceptance of the recommendation [18, 54].

H1: Questioning GenAI style elicits higher cognitive control/engagement during the decision episode than confident style.

TTD posits that decision aids influence decisions by becoming dominant inputs to judgment, increasing reliance [50, 51, 55]. Decision-support research shows substantial reliance of algorithmic advice even when its logic is opaque (i.e., algorithm appreciation) [53, 58, 59]. Accordingly, GenAI support should increase reliance relative to baseline. SIT further implies that high certainty and normative force amplify reliance, whereas questioning prompts reduce it by inducing scrutiny [54].

H2: GenAI assistance leads to higher reliance on the system's output than baseline.

H3: Reliance is higher under the confident style than under the questioning style.

Beyond advice framing, reliance depends on how actively decision makers evaluate the recommendation. Higher cognitive control and engagement should increase scrutiny and reduce uncritical acceptance. This aligns with TTD's view of reliance as integrating the aid into judgment (not by mere exposure) [52, 55], arguing for NeuroIS perspectives that process measures at decision time beyond post-hoc rationales [19, 57].

H4: Higher cognitive control/engagement during the decision episode is associated with lower reliance on GenAI output.

In high-stakes decisions, reliance may blur accountability by shifting perceived ownership away from the decision maker [60, 61]. Consistent with responsibility diffusion and TTD, greater dominance of GenAI output should reduce perceived authorship, accountability, and agency [14, 50, 62]. Thus, reliance should mediate effects of GenAI availability and style on accountability: assistance and high–normative-force advice increase reliance, which reduces perceived accountability.

H5: Higher reliance on GenAI is associated with lower perceived accountability.

H6: The effects of (a) GenAI assistance (vs. baseline) and (b) GenAI style on accountability are mediated by reliance.

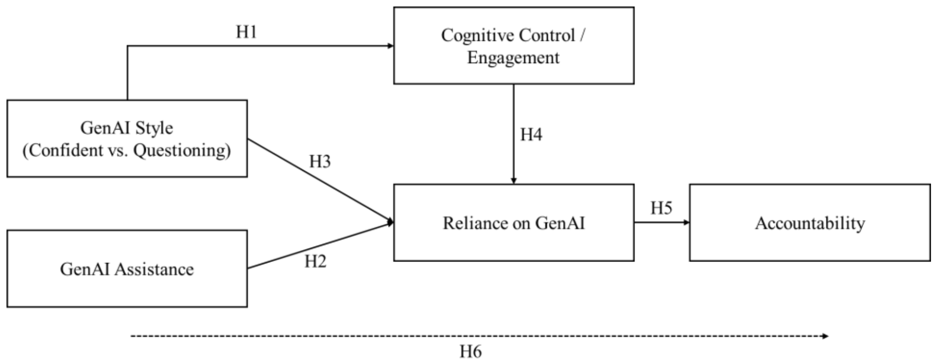


Fig. 4. Research model

Methodology

Study Design: We propose a controlled within-subject laboratory experiment to examine how GenAI assistance and GenAI interaction style shape (i) cognitive control/engagement, (ii) reliance on GenAI, and (iii) downstream accountability/ownership perceptions in a recruitment tie-breaker task. Participants complete three randomized blocks (two decisions each; six total): baseline (*no GenAI*), GenAI with a *confident high-normative-force* style, and GenAI with a *questioning reflection-fostering style*. The core decision setting is designed as a tie-breaker between two equally suitable candidates, such that the experimental manipulation targets advice integration and justification under ambiguity rather than simple performance differences between applicants.

Task and Experimental Manipulation: Participants act as a decision maker selecting one of two equally fitting applicants for a student assistant job, ensuring task accessibility for a student sample (see *Sample and Procedure*) while retaining reliance and accountability trade-offs. The candidate profiles are constructed so that both applicants appear comparably suitable overall but differ on selected attributes that make the final choice defensible in more than one way. This creates the intended tie-breaker situation for decision-making and increases the relevance of GenAI-generated rationales. In GenAI blocks, participants may consult the assistant before committing (cf. Fig. 5); the system provides assistance, recommendation, and rationale with style manipulated as follows: The *confident high-normative-force* style uses high decisional certainty and prescriptive language (e.g., stronger 'should' statements), whereas the *questioning reflection-fostering* style foregrounds uncertainty, reflection, and prompts for justification. The system utilizes the OpenAI API with the gpt-5.4 model and respective system prompts to ensure the defined interaction styles.

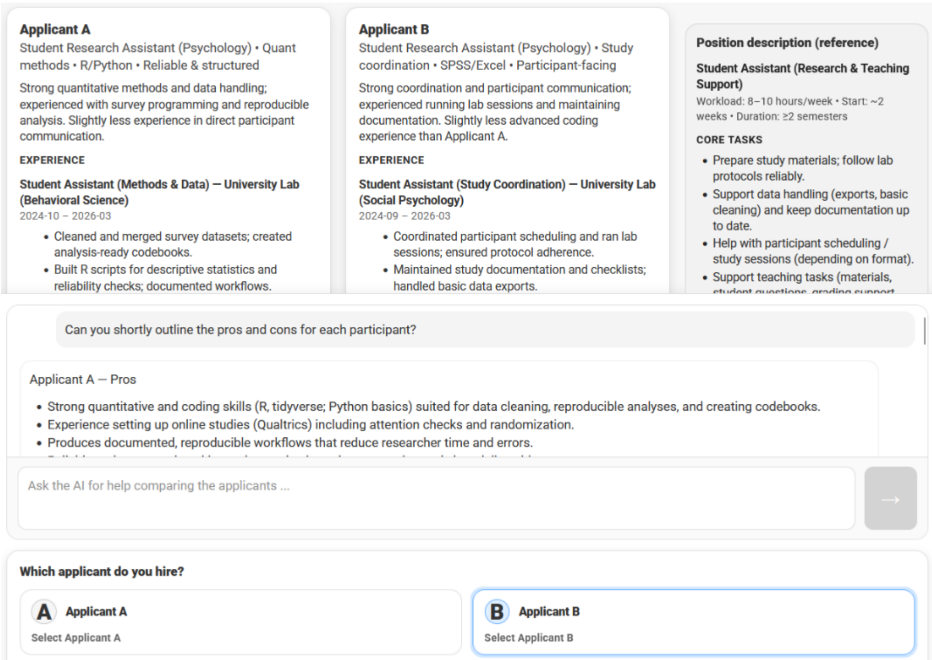


Fig. 5. Screenshot of the genAI-supported decision task

Sample and Procedure: Participants will be recruited from a German student pool. A sample size of n = 50 is determined via a priori power analysis in G*Power (α = .05, power $\geq$ .80, effect size = .2) which suggests a minimum of 42 participants. After consent and demographics, participants are equipped with a Unicorn Hybrid Black EEG and calibrated on a Tobii Pro Spark eye tracker. As outlined in Fig. 6, participants complete a practice trial, resting phases for EEG calibration, and then proceed through the three randomized blocks (two hiring decisions per treatment). To address EEG fatigue effects, we employ resting phases before each task block. Before each decision, participants are first presented with the hiring scenario. They then view the two candidate profiles simultaneously on the same page, together with the job description and the final choice options. Each decision is limited to 5 minutes. As shown in Fig. 5, in the GenAI conditions, the same decision page additionally contains the GenAI interaction area with advice. Thus, each trial is centered on a single integrated comparison screen rather than on separate sequential profile pages. This setup is intended to resemble a realistic tie-breaker comparison in which decision makers directly contrast two similarly suitable applicants while optionally consulting GenAI support. To increase consequentiality and accountability salience without real economic risk, participants report a mock 'stake' of an intended performance bonus on each decision as a proxy for commitment. The bonus is always paid in full; the stake is incentive-mimicking only and serves as a continuous endorsement measure to facilitate faithful decision-making.

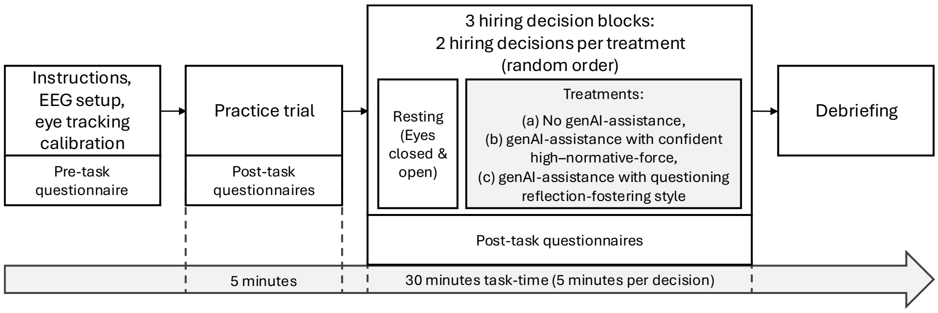


Fig. 6. Procedure of the experiment

Measures: For each decision, we record (i) final choice, (ii) decision time, (iii) interaction traces in GenAI blocks (e.g., prompt count, total interaction time, and complete chat logs), and (iv) the mock stake. Behavioral reliance is captured via agreement/override rates (i.e., match to recommendation). Self-reports include TTD-aligned reliance [55], algorithm aversion/appreciation [58], cognitive absorption [63], decision difficulty [64], and task-adapted sense of agency as an accountability measure [65]. Brief manipulation checks assess perceived decisional certainty/normative force versus reflection prompting in the respective GenAI style blocks. EEG is recorded to capture objective, in-task indicators of the focal constructs while participants process candidate information, consult GenAI advice, and commit to a final choice. Rather than treating EEG as a separate outcome domain, we use it as a process-tracing measurement approach for cognitive control and engagement during the decision episode. We record EEG data using a reduced 10-20 montage comprising Fz, F3, F4, Cz, P3, Pz, and P4, with TP10 recorded as the mastoid reference channel for re-referencing. This montage is selected to capture the frontal-midline and posterior scalp regions most relevant to the focal constructs. In line with prior EEG research, cognitive control is operationalized primarily through frontal-midline theta activity over Fz and Cz, where theta-band dynamics are commonly linked to conflict monitoring and control processes [66], whereas attentional engagement is operationalized through posterior alpha power and alpha suppression over P3, Pz, and P4, where alpha dynamics have been associated with visual attention and task engagement [67, 68]. The additional frontal electrodes F3 and F4 are included to broaden coverage of frontal scalp dynamics beyond the frontal-midline sites. Although the primary theta analyses remain centered on Fz and Cz, the inclusion of F3 and F4 preserves a compact frontal cluster that can support data quality assessment and exploratory analyses of broader frontal scalp-level activation patterns.

We distinguish between task-level and behavior-defined EEG analyses. First, we examine the overall decision episode from stimulus onset to final commitment. Second, we use eye-tracking data to define more fine-grained EEG windows based on participants' visual attention to specific areas of interest. More specifically, area of interest (AOI) fixations are used to segment EEG activity into behavior-defined windows associated with processing candidate information versus processing GenAI output (e.g., recommendation reading versus CV reading). This approach allows us to relate oscillatory dynamics more directly to the information currently attended to, thereby differentiating general task-related engagement from more specific neural dynamics during information acquisition and advice evaluation. Eye movements are recorded in parallel and analyzed using AOI-based measures of evidence acquisition and reliance, including dwell-time proportions on candidate and GenAI AOIs, switching behavior between AOIs, and post-advice revisits to candidate information for verification [23, 69].

Data Analysis: Data will be analyzed using mixed-effects models to account for repeated decisions and trial heterogeneity, estimating condition effects on behavioral, self-report, eye-tracking, and EEG outcomes, and testing whether attentional and neurocognitive measures are associated with responsibility-related responses. The mixed-effects specification is particularly suitable because it models repeated tie-breaker decisions nested within participants while accounting for trial-level heterogeneity in candidate constellations and advice conditions.

EEG preprocessing and spectral quantification are conducted in Python using the MNE ecosystem prior to model estimation [70]. The continuous EEG is re-referenced to a linked-mastoid reference based on the right mastoid channel (TP10). Preprocessing is conducted separately for each task window. First, bad channels are identified based on deviation, high-frequency noise, and flatness criteria and are subsequently interpolated. Second, Artifact Subspace Reconstruction (ASR) is applied to attenuate high-amplitude transient artifacts while preserving the task-related signal [71]. The cleaned data are then segmented into 2 second epochs with 50% overlap.

Spectral quantification is based on multitaper power spectral density (PSD) estimation, followed by averaging across the channels relevant to the respective construct. To derive individualized frequency bands, resting-state and task-related spectral estimates are first computed at the participant level [72]. Individual alpha frequency is then determined in a multi-step procedure [73]. First, the aperiodic 1/f component is fitted and subtracted from the mean PSD. Second, the flattened alpha-band curve is smoothed. Third, the most prominent local peak in the 7-13 Hz range is selected. If this peak is located at the edge of the search window, or if no peak exceeds the required prominence threshold, the procedure falls back to a local center-of-gravity estimate based on the raw PSD. The individualized alpha band is then defined as IAF ± 1.5 Hz.

To derive individualized theta-band quantification, the theta–alpha transition frequency is estimated following Klimesch's approach as the intersection of resting-state and task-related PSD curves [74]. Based on this transition estimate, the theta band is defined as the 2 Hz range showing the largest difference between resting and working spectra. This individualized band definition is intended to account for interindividual variation in oscillatory peak structure and thereby improve the construct validity of the EEG measures.

Outlook

This study proposal advances NeuroIS research on human-GenAI collaboration by testing a theory-grounded mechanism model in a high-stakes, fairness-sensitive decision context. The planned experiment will provide convergent behavioral, self-report, EEG, and eye-tracking evidence on how GenAI interaction style shapes cognitive control/engagement, reliance, and perceived accountability in hiring tie-breaking decision scenarios. First study data points for a pre-study will be collected and analyzed for discussion at the NeuroIS retreat. Beyond contributing to Technology Dominance Theory in the GenAI era, the results can inform the design and governance of responsible GenAI decision support by identifying interaction styles that calibrate reliance without undermining accountability.

References

1. Banh, L., Strobel, G.: Generative artificial intelligence. Electron Markets, vol. 33 (2023). doi: 10.1007/s12525-023-00680-1
2. Seeber, I., Bittner, E., Briggs, R.O., Vreede, T. de, Vreede, G.-J. de, Elkins, A., Maier, R., Merz, A.B., Oeste-Reiß, S., Randrup, N., et al.: Machines as teammates: A research agenda on AI in team collaboration. Information & Management, vol. 57, 103174 (2020). doi: 10.1016/j.im.2019.103174
3. Boussioux, L., Lane, J.N., Zhang, M., Jacimovic, V., Lakhani, K.R.: The Crowdless Future? Generative AI and Creative Problem-Solving. Org. Sci., vol. 35, 1589–1607 (2024). doi: 10.1287/orsc.2023.18430
4. Schulz, T., Speck, C.: Conceptualising Reflective Use: Toward a Process Perspective on Human-AI Interaction. Proceedings of the European Conference on Information Systems 2026, vol. Forthcoming (2026)
5. Anthony, C., Bechky, B.A., Fayard, A.-L.: "Collaborating" with AI: Taking a System View to Explore the Future of Work. Organization Science, vol. 34, 1672–1694 (2023). doi: 10.1287/orsc.2022.1651
6. Krakowski, S.: Human-AI agency in the age of generative AI. Information and Organization, vol. 35, 100560 (2025). doi: 10.1016/j.infoandorg.2025.100560
7. Holldack, F., Banh, L., Strobel, G.: Agentic information systems. Electron Markets, vol. 36 (2026). doi: 10.1007/s12525-025-00861-0
8. Fabre, E.F., Mouratille, D., Bonnemains, V., Palmiotti, G.P., Causse, M.: Making moral decisions with artificial agents as advisors. A fNIRS study. Computers in Human Behavior: Artificial Humans, vol. 2, 100096 (2024). doi: 10.1016/j.chbah.2024.100096
9. Bertheau, M.A.K., Boetzel, C., Herrmann, C.S.: Event-related potentials reveal incongruent behavior of autonomous vehicles in the moral machine dilemma. Scientific reports, vol. 15, 16048 (2025). doi: 10.1038/s41598-025-99531-6
10. Chen, Z.: Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. Cognition, technology & work (Online), vol. 25, 135–149 (2023). doi: 10.1007/s10111-022-00716-0

11. Pan, Y., Froese, F.J.: An interdisciplinary review of AI and HRM: Challenges and future directions. Human Resource Management Review, vol. 33, 100924 (2023). doi: 10.1016/j.hrmr.2022.100924
12. Banh, L., Rex, A., Strobel, G., Urbach, N.: Hiring Tomorrow's Talents: How Generative Artificial Intelligence Transforms Human Resources Recruitment. In: Bui, T.X. (ed.) Proceedings of the 59th Hawaii International Conference on System Sciences. Honolulu, HI (2026)
13. Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G.J., Beltran, J.R., Boselie, P., Lee Cooke, F., Decker, S., DeNisi, A., et al.: Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. Human Res Mgmt Journal, vol. 33, 606–659 (2023). doi: 10.1111/1748-8583.12524
14. Salatino, A., Prével, A., Caspar, E., Lo Bue, S.: Influence of AI behavior on human moral decisions, agency, and responsibility. Scientific reports, vol. 15, 12329 (2025). doi: 10.1038/s41598-025-95587-6
15. Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.-H., Beresnitzky, A.V., Braunstein, I., Maes, P.: Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv (2025)
16. Wang, Y., Xu, M., Zhao, L.: Exploring the Cognitive Mechanisms Behind the Adoption of Algorithmic Advice. Bus Inf Syst Eng, vol. (2025). doi: 10.1007/s12599-025-00925-7
17. León-Domínguez, U.: Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions. Neuropsychology, vol. 38, 293–308 (2024). doi: 10.1037/neu0000948
18. Dimoka, A., Davis, F.D., Gupta, Pavlou, Banker, R.D., Dennis, A.R., Ischebeck, Müller-Putz, G.R., Benbasat, I., Gefen, et al.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly, vol. 36, 679 (2012). doi: 10.2307/41703475
19. Riedl, R., Davis, F.D., Hevner, A.: Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement. Journal of the Association for Information Systems, vol. 15, I–XXXV (2014). doi: 10.17705/1jais.00377
20. Fronda, G., Angioletti, L., Balconi, M.: EEG Correlates of Moral Decision-Making: Effect of Choices and Offers Types. AJOB neuroscience, vol. 15, 191–205 (2024). doi: 10.1080/21507740.2024.2306270
21. Garrigan, B., Adlam, A.L.R., Langdon, P.E.: The neural correlates of moral decision-making: A systematic review and meta-analysis of moral evaluations and response decision judgements. Brain and cognition, vol. 108, 88–97 (2016). doi: 10.1016/j.bandc.2016.07.007
22. Weber, B., Fischer, T., Riedl, R.: Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review. Journal of Systems and Software, vol. 178, 110946 (2021). doi: 10.1016/j.jss.2021.110946
23. Pina, A., Petersheim, C., Cherian, J., Lahey, J.N., Alexander, G., Hammond, T.: Using Machine Learning with Eye-Tracking Data to Predict if a Recruiter Will Approve a Resume. MAKE, vol. 5, 713–724 (2023). doi: 10.3390/make5030038

24. Herath Pathirannehelage, S., Shrestha, Y.R., Krogh, G. von: Design principles for artificial intelligence-augmented decision making: An action design research study. EJIS, vol. 34, 207–229 (2025). doi: 10.1080/0960085X.2024.2330402
25. Cao, S., Huang, C.-M.: Understanding User Reliance on AI in Assisted Decision-Making. Proc. ACM Hum.-Comput. Interact., vol. 6, 1–23 (2022). doi: 10.1145/3555572
26. Jussupow, E., Spohrer, K., Heinzl, A., Gawlitza, J.: Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. Information Systems Research, vol. 32, 713–735 (2021). doi: 10.1287/isre.2020.0980
27. Lebovitz, S., Levina, N., Lifshitz-Assaf, H.: Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. MIS Quarterly, vol. 45, 1501–1526 (2021). doi: 10.25300/MISQ/2021/16564
28. Pieper, M., Gleasure, R.: How AI Helps to Compile Human Intelligence: An Empirical Study of Emerging Augmented Intelligence for Medical Image Scanning. Information Systems Journal, vol. 35, 1399–1421 (2025). doi: 10.1111/isj.12585
29. Wang, W., Gao, G., Agarwal, R.: Friend or Foe? Teaming Between Artificial Intelligence and Workers with Variation in Experience. Mang. Sci., vol. (2023). doi: 10.1287/mnsc.2021.00588
30. Vössing, M., Kühl, N., Lind, M., Satzger, G.: Designing Transparency for Effective Human-AI Collaboration. Information Systems Frontiers, vol. 24, 877–895 (2022). doi: 10.1007/s10796-022-10284-3
31. Köbis, N., Rahwan, Z., Rilla, R., Supriyatno, B.I., Bersch, C., Ajaj, T., Bonnefon, J.-F., Rahwan, I.: Delegation to artificial intelligence can increase dishonest behaviour. Nature, vol. 646, 126–134 (2025). doi: 10.1038/s41586-025-09505-x
32. Bauer, K., Gill, A.: Mirror, Mirror on the Wall: Algorithmic Assessments, Transparency, and Self-Fulfilling Prophecies. Information Systems Research, vol. 35, 226–248 (2024). doi: 10.1287/isre.2023.1217
33. Ghasemaghaei, M., Kordzadeh, N.: Ethics in the Age of Algorithms: Unravelling the Impact of Algorithmic Unfairness on Data Analytics Recommendation Acceptance. Information Systems Journal, vol. 35, 1166–1197 (2025). doi: 10.1111/isj.12572
34. Köbis, N., Bonnefon, J.-F., Rahwan, I.: Bad machines corrupt good morals. Nature Human Behaviour, vol. 5, 679–685 (2021). doi: 10.1038/s41562-021-01128-2
35. Polyportis, A., Pahos, N.: Understanding students' adoption of the ChatGPT chatbot in higher education: the role of anthropomorphism, trust, design novelty and institutional policy. Behaviour & Information Technology, vol. 44, 315–336 (2025). doi: 10.1080/0144929X.2024.2317364
36. Huynh, M.-T.: Using generative AI as decision-support tools: unraveling users' trust and AI appreciation. Journal of Decision Systems, vol. , 1–32 (2024). doi: 10.1080/12460125.2024.2428166

37. Christensen, J.F., Gomila, A.: Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. Neuroscience and biobehavioral reviews, vol. 36, 1249–1264 (2012). doi: 10.1016/j.neubiorev.2012.02.008
38. Banh, L., Stangl, F.J., Strobel, G., Riedl, R.: Exploring the NeuroIS Potential for Generative Artificial Intelligence: Findings from a Literature Review. Proceedings of the 17th NeuroIS Retreat, vol. (2025)
39. Grange, C., Demazure, T., Ringeval, M., Bourdeau, S., Martineau, C.: The Human-GenAI Value Loop in Human-Centered Innovation: Beyond the Magical Narrative. Information Systems Journal, vol. 36, 29–51 (2026). doi: 10.1111/isj.12602
40. Tadson, B., Krisam, C., Mädche, A.: Understanding Overreliance in Human-GenAI Interactions and Its Impact on Divergent Thinking: A NeuroIS Study. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., B. Randolph, A., R. Müller-Putz, G. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 9, pp. 143–157. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-032-00815-2_14
41. Chen, Z., Chan, J.: Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise. Mang. Sci., vol. 70, 9101–9117 (2024). doi: 10.1287/mnsc.2023.03014
42. Doshi, A.R., Hauser, O.P.: Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science advances, vol. 10, eadn5290 (2024). doi: 10.1126/sciadv.adn5290
43. Schulz, T., Krisam, C., Seitz, J.: EyeGPT: A Cognitive Load-Adaptive GenAI Assistant with Eye Tracking for Programming Education. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., B. Randolph, A., R. Müller-Putz, G. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 9, pp. 45–55. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-032-00815-2_4
44. Conrad, C., Boakye, N., Newman, A.J.: Do We Perceive Virtual Teachers as Human? An EEG Experiment Proposal for Investigating the Cognition of AI Teaching Expectations. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., B. Randolph, A., R. Müller-Putz, G. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 9, pp. 57–64. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-032-00815-2_5
45. Knierim, M.T., Del Puppo, L.: Do Generative AI Tools Foster Positive Experiences in Knowledge Work? A NeuroIS Research Proposal. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G. (eds.) Proceedings NeuroIS Retreat 2023, pp. 53–61 (2023)
46. Ritz, E., Freise, L.R., Li, M.M.: Offloading to Digital Minds: How Generative AI Can Help to Craft Jobs. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G. (eds.) Proceedings NeuroIS Retreat 2024, pp. 9–18 (2024)
47. Benbya, H., Strich, F., Tamm, T.: Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. JAIS, vol. 25, 23–36 (2024). doi: 10.17705/1jais.00861

48. Schulz, T., Knierim, M.: Cognitive Load Dynamics in Generative AI-Assistance: A NeuroIS Study. Proceedings of the International Conference on Information Systems 2024, vol. (2024)
49. Schulz, T., Knierim, M.T., Weinhardt, C.: How Generative-AI-Assistance Impacts Cognitive Load During Knowledge Work: A Study Proposal. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G.R. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 66, pp. 357–365. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-031-71385-9_31
50. Sutton, S.G., Arnold, V., Holt, M.: An extension of the theory of technology dominance: Capturing the underlying causal complexity. International Journal of Accounting Information Systems, vol. 50, 100626 (2023). doi: 10.1016/j.accinf.2023.100626
51. Ebrahimi, S., Matt, C.: Not seeing the (moral) forest for the trees? How task complexity and employees' expertise affect moral disengagement with discriminatory data analytics recommendations. Journal of Information Technology, vol. 39, 477–502 (2024). doi: 10.1177/02683962231181148
52. Chang, T.-S., Huang, L.-G.: The impact of generative artificial intelligence on software development performance: a study based on the theory of technology dominance. Information & Management, vol. 63, 104306 (2026). doi: 10.1016/j.im.2026.104306
53. Logg, J.M., Minson, J.A., Moore, D.A.: Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, vol. 151, 90–103 (2019). doi: 10.1016/j.obhdp.2018.12.005
54. Kelman, H.C.: Compliance, identification, and internalization three processes of attitude change. Journal of Conflict Resolution, vol. 2, 51–60 (1958). doi: 10.1177/002200275800200106
55. Hampton, C.: Determinants of reliance: An empirical test of the theory of technology dominance. International Journal of Accounting Information Systems, vol. 6, 217–240 (2005). doi: 10.1016/j.accinf.2005.10.001
56. Kirwan, C.B., Vance, A., Jenkins, J.L., Anderson, B.B.: Embracing brain and behaviour: Designing programs of complementary neurophysiological and behavioural studies. Information Systems Journal, vol. 33, 324–349 (2023). doi: 10.1111/isj.12402
57. Dimoka, A., Pavlou, P.A., Davis, F.D.: Research Commentary —NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. Information Systems Research, vol. 22, 687–702 (2011). doi: 10.1287/isre.1100.0284
58. Jussupow, E., Benbasat, I., Heinzl, A.: An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making. MIS Quarterly, vol. 48, 1575–1590 (2024). doi: 10.25300/MISQ/2024/18512
59. You, S., Yang, C.L., Li, X.: Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation? Journal of Management Information Systems, vol. 39, 336–365 (2022). doi: 10.1080/07421222.2022.2063553

60. Cooper, A.F., Moss, E., Laufer, B., Nissenbaum, H.: Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In: 2022 ACM Conference on Fairness Accountability and Transparency, pp. 864–876. ACM, New York, NY, USA (2022). doi: 10.1145/3531146.3533150
61. Besio, C., Fedtke, C., Grothe-Hammer, M., Karafillidis, A., Pronzini, A.: Algorithmic responsibility without accountability: Understanding data-intensive algorithms and decisions in organisations. Syst Res Behav Sci, vol. 42, 739–755 (2025). doi: 10.1002/sres.3028
62. Li, P., Jia, S., Feng, T., Liu, Q., Suo, T., Li, H.: The influence of the diffusion of responsibility effect on outcome evaluations: electrophysiological evidence from an ERP study. NeuroImage, vol. 52, 1727–1733 (2010). doi: 10.1016/j.neuroimage.2010.04.275
63. Agarwal, R., Karahanna, E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage1. MIS Quarterly, vol. 24, 665–694 (2000). doi: 10.2307/3250951
64. Zhang, Y., Mittal, V.: Decision Difficulty: Effects of Procedural and Outcome Accountability. J CONSUM RES, vol. 32, 465–472 (2005). doi: 10.1086/497558
65. Tapal, A., Oren, E., Dar, R., Eitam, B.: The Sense of Agency Scale: A Measure of Consciously Perceived Control over One's Mind, Body, and the Immediate Environment. Frontiers in psychology, vol. 8, 1552 (2017). doi: 10.3389/fpsyg.2017.01552
66. Duprez, J., Gulbinaite, R., Cohen, M.X.: Midfrontal theta phase coordinates behaviorally relevant brain computations during cognitive control. NeuroImage, vol. 207, 116340 (2020). doi: 10.1016/j.neuroimage.2019.116340
67. Magosso, E., Crescenzio, F. de, Ricci, G., Piastra, S., Ursino, M.: EEG Alpha Power Is Modulated by Attentional Changes during Cognitive Tasks and Virtual Reality Immersion. Computational intelligence and neuroscience, vol. 2019, 7051079 (2019). doi: 10.1155/2019/7051079
68. Songlin, Q., Xia, X., Chen, J., Valsecchi, M.: Attentional tracking reduces cortical alpha oscillations. Scientific reports, vol. 15, 34067 (2025). doi: 10.1038/s41598-025-14585-w
69. Khalili, N., Ma, C., Xu, J., Lauwereyns, J.: Exploring gaze dynamics in moral decision making: Extending the gaze cascade effect. In: Sugano, Y., Khamis, M., Chetouani, A., Sidenmark, L., Bruno, A. (eds.) Proceedings of the 2025 Symposium on Eye Tracking Research and Applications, pp. 1–3. ACM, New York, NY, USA (2025). doi: 10.1145/3715669.3726807
70. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al.: MEG and EEG data analysis with MNE-Python. Frontiers in neuroscience, vol. 7, 267 (2013). doi: 10.3389/fnins.2013.00267
71. Miyakoshi, M.: Artifact subspace reconstruction: a candidate for a dream solution for EEG studies, sleep or awake. Sleep, vol. 46 (2023). doi: 10.1093/sleep/zsad241

72. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain research. Brain research reviews, vol. 29, 169–195 (1999). doi: 10.1016/S0165-0173(98)00056-3
73. Goljahani, A., D'Avanzo, C., Schiff, S., Amodio, P., Bisiacchi, P., Sparacino, G.: A novel method for the determination of the EEG individual alpha frequency. NeuroImage, vol. 60, 774–786 (2012). doi: 10.1016/j.neuroimage.2011.12.001
74. Vallarino, E., Sommariva, S., Famà, F., Piana, M., Nobili, F., Arnaldi, D.: Transfreq: A Python package for computing the theta-to-alpha transition frequency from resting state electroencephalographic data. Human brain mapping, vol. 43, 5095–5110 (2022). doi: 10.1002/hbm.25995

Beyond Single Metrics: Validating Composite Indicators for Physiological and Cognitive User Experience Research

Marcin Czaban✉[1], Joachim Riedl[2], Stefan Wengler[2]

[1] Betriebswirtschaftliches Forschungszentrum für Fragen der mittelständischen Wirtschaft e.V., Bayreuth, Germany
marcin.czaban@bfm-bayreuth.de
[2] University of Applied Sciences Hof, Faculty of Economics, Hof, Germany
{joachim.riedl; stefan.wengler}@hof-university.de

Abstract. In NeuroIS research, the triangulation of physiological data and self-reported measures often leads to inconsistent results, complicating the assessment of user activation and stress. This study addresses this reliability gap by evaluating the suitability of composite indicators versus single measurements. In a driving simulator study (n=68) comparing real-world and simulated driving across seven segments, we collected Galvanic Skin Response (GSR), Heart Rate (HR), salivary cortisol, and cognitive workload data (NASA-TLX, single-item stress). Principal component analysis revealed two stable dimensions: Physiological Reaction (PR) and Cognitive Reaction (CR). Validation using salivary cortisol levels indicates that these composite indicators offer higher stability and explanatory power than single metrics. We propose these composite indicators as a robust methodological approach for future NeuroIS studies to mitigate artifacts and measurement divergence.

Keywords: composite indicators· stress measurement · technostress · multimodal measurement

Introduction

A core objective of NeuroIS is to employ neurophysiological tools to better understand information systems [1]. However, a central challenge, particularly in the study of technostress, is the "triangulation problem": physiological data (e.g., GSR, HR) and subjective self-reports often yield heterogeneous or conflicting results regarding user stress [2, 3]. While individual findings suggest that task characteristics correlate with physiological response, the inconsistency of single measurements undermines the validity of User Experience findings. In applied contexts, researchers often rely on single indicators (e.g., HR), which are prone to artifacts such as movement or environmental noise [4]. This heterogeneity limits practical applicability when results hinge on a specific indicator [5].

This paper addresses this gap by proposing composite indicators that aggregate specific physiological and cognitive measures into potentially coherent higher-order dimensions. Building on the call for “measurement pluralism” [6], we pose the research question: Can such composite indicators provide a more stable and valid assessment of user stress reactions than single measurements? We validate these indicators using salivary cortisol, a direct biological marker of hypothalamic-pituitary-adrenal (HPA) axis activity [6].

Theoretical Background and Hypotheses

Stress Measurement and Artifacts

Technostress is defined as "the stress experienced by people as a result of their interactions with technologies" [7]. While traditional driving research focuses on traffic frustration, modern vehicles are highly complex IT artifacts. Drivers continuously interact with sophisticated Human-Machine Interfaces (HMIs) and digital assistance systems. Consequently, driving has become a demanding human-computer interaction (HCI) task where information overload can easily exceed user capabilities, triggering technostress [8]. o measure this arousal, NeuroIS researchers utilize Autonomic Nervous System (ANS) responses via, e.g., Galvanic Skin Response (GSR) and Electrocardiography (ECG) [9]. However, widely used metrics like Heart Rate (HR) and Skin Conductance Response (SCR) are sensitive to artifacts and may reflect unspecific arousal rather than negative stress [10]. Consequently, stress is a multifaceted phenomenon involving both autonomic activation and cognitive appraisal [11] that cannot be reliably inferred from a single signal.

Composite Indicators & Cortisol Validation

Following latent variable modeling principles, we hypothesize that stress states manifest across multiple channels simultaneously [3]. By aggregating these into a Physiological Reaction (PR) and a Cognitive Reaction (CR), we aim to reduce error variance inherent in single-sensor measurements [12]. Such aggregation can enhance stability by capturing shared variance while reducing indicator-specific noise [13]. To evaluate whether measures reflect genuine stress, cortisol serves as a biological criterion. Unlike fast-reacting ANS measures, cortisol reflects a cumulative hormonal response to stress [14, 15]. In the present design, cortisol is used to validate the stress relevance of physiological single measures and the derived composite indicators.

Based on this reasoning, we propose the following hypotheses: **H1 (Factor Structure Stability):** The two-dimensional factor structure of the composite stress indicators (Physiological Reaction and Cognitive Reaction) exhibits measurement stability across varying situational driving demands. **H2 (Load-Response Relationship):** There is a significant inverse relationship between a participant's perceived ability to manage the situational load and the magnitude of both composite indicators (PR and CR). **H3 (Biological Validation):** Under conditions of high situational demand, both the Physiological Reaction (PR) and Cognitive Reaction (CR) composite indicators exhibit a significant positive correlation with the biological stress marker salivary cortisol.

Methodology

Participants and Design

We conducted a study with N=68 participants (37 females, 31 males, M age =30.07, SD=11.58). The underlaying data have already been examined with respect to simulator validity in terms of physiological responses [16, 17], but not with regard to indicator aggregation.

To encompass a broad range of situational demands, the within-subject design compared two experimental conditions: a real-world drive and a simulated drive. The task involved navigating a 23 km circular route, explicitly designed to induce varying levels of workload. It consisted of seven distinct segments: (1) City/outskirts, (2) Village, (3) Rural road to motorway entrance, (4) Motorway, (5) Motorway exit, (6) Village/rural road, and (7) City traffic, yielding reaction data for 14 unique conditions per participant.

The real-world drive was conducted in regular traffic under favorable weather conditions using a standard rental car (110 kW). To ensure exact comparability, the real-world route was digitally replicated as a 1:1 scale "digital twin" on a medium-fidelity simulator [18]. The simulator was operated via Silab 7.1 software (WIVW) and featured an original driver's seat, a force-feedback steering wheel, a 180° horizontal field of view via LCD screens, and a 3-degree-of-freedom D-Box to provide physical feedback on road surface conditions.

The procedure followed a strict, standardized protocol: After an initial baseline salivary cortisol test (SCT/0) and a pre-questionnaire to assess baseline stress/cognitive load, participants were equipped with the physiological sensors (GSR, ECG). To establish a baseline and avoid simulator sickness artifacts during the initial phase, participants always completed the real-world drive first. This was immediately followed by the administration of the cognitive questionnaires and a second cortisol sample (SCT/1). Subsequently, participants were introduced to the simulator via three standardized training routes to prevent habituation issues. After completing the simulated drive on the digital twin route, participants answered the identical set of questionnaires a final time, followed by the concluding cortisol test (SCT/2).

Measurements

Physiological. GSR (Shimmer 3 GSR+) to extract Skin Conductance Response (SCR) and Level (SCL) [9]; ECG to record HR and HRV indices. Data were synchronized via iMotions 9.1.

Cognitive. Workload via NASA-TLX ($\alpha > .72$) [19]; Stress via Short Stress State Questionnaire (SSSQ) [20] and a single Item (Stress) ; Physical Wellbeing (PW) (reverse-coded) [21]; Situational load via Vehicle Operation (VO) (self-developed single-item).

Validation. Salivary Cortisol (SCT) samples taken pre- and post-drive, analyzed via chemiluminescence immunoassay.

Data Analysis. Data were restructured so that each participant × segment constituted an observation (n=476 per condition) to retain variance. Composite indicators were derived using principal component analysis (PCA) with varimax rotation.

Results

Descriptive Findings and Manipulation check

First, we compared the two conditions to check the manipulation. The simulator drive induced significantly higher loads than the real-world drive across almost all metrics. NASA-TLX was significantly higher in the simulator (M=4.86) compared to the real drive (M=2.63, $p<.001$). Similarly, single-item Stress was higher in the simulator (5.63 vs. 2.84, $p<.01$).

Interestingly, physiological metrics followed this trend: SCL and peak amplitude were significantly lower in the real drive. Cortisol levels also showed a trend of increase in the simulator condition, contrasting with the natural diurnal decline observed during the real drive. In total, nine of 12 indicators suggested that the simulator drive was significantly more stressful than the real-world drive. This variation is essential for our methodological approach, as it enables us to examine driving situations with differing stress intensities. Consequently, the reliability of the composite indicators can be analyzed independent of the specific stress level.

Derivation and Stability of Composite Indicators (Testing H1)

To identify the most reliable single indicators, we conducted an initial correlation analysis, which showed that individual indicators provided parallel but not fully consistent signals. We selected indicators for aggregation based on evidence: SCR was retained over SCL because it correlated on a higher level with cognitive indicators; HR was retained as the most robust cardiac measure (reflecting findings by [22]). To identify the underlying factor structure, we initially conducted a Principal Component Analysis (PCA) with varimax rotation across the entire driving dataset (real and simulated conditions).

Table 1 Varimax Rotated Factor Loadings of Calculated Composite Indicators

	Real Ride Reaction Data		Simulator Ride Reaction Data	
Items	Cognitive	Physiological	Cognitive	Physiological
PW	**.824**	.147	**.839**	.158
NASA-TLX	**.869**	-.036	**.931**	.017
Stress	**.850**	-.093	**.911**	.063
SCR	.309	**.757**	.136	**.748**
HR	-.280	**.750**	.002	**.795**

As Table 1 illustrates, this yielded a consistent two-factor solution: the physiological variables (HR, SCR) formed a distinct factor, termed Physiological Reaction (PR), while the subjective measures (NASA-TLX, perceived stress, physical wellbeing) reliably clustered into a second factor, termed Cognitive Reaction (CR).

Having established these composite indicators, we tested their measurement invariance across varying situational driving demands (H1). To this end, separate PCAs were calculated for all 14 driving segments.

Table 2 PCA Rotated Factor Loading across the Seven Sections

Section	Indicator	Real Ride		Simulator Ride	
		CR	PR	CR	PR
1: City, city outskirts, rural road	PW	**.834**	.206	**.834**	.155
	NASA TLX	**.865**	-.107	**.934**	-.004
	STRESS	**.856**	-.105	**.906**	.117
	SCR	.220	**.811**	.207	**.735**
	HR	-.229	**.715**	-.039	**.837**
	KMO	.660		.700	
	Bartlett	<.001		<.001	
	Variance explained	70.277		74.224	
2: Village	PW	**.827**	.153	**.842**	.131
	NASA TLX	**.864**	-.077	**.928**	.051
	STRESS	**.834**	-.097	**.913**	.041
	SCR	.444	**.625**	.116	**.786**
	HR	-.265	**.799**	.018	**.806**
	KMO	.715		.700	
	Bartlett	<.001		<.001	
	Variance explained	69.219		74.137	
3: Rural road to motorway entrance	PW	**.822**	.159	**.845**	.117
	NASA TLX	**.862**	-.061	**.930**	.002
	STRESS	**.857**	-.071	**.911**	.052
	SCR	.339	**.751**	.103	**.766**
	HR	-.294	**.768**	-.003	**.788**
	KMO	.702		.698	
	Bartlett	<.001		<.001	
	Variance explained	70.847		72.935	
4: Motorway	PW	**.819**	.130	**.836**	.156
	NASA TLX	**.874**	-.002	**.932**	.030
	STRESS	**.847**	-.058	**.902**	.126
	SCR	.409	**.711**	.141	**.464**
	HR	-.234	**.839**	-.025	**.901**
	KMO	.702		.700	
	Bartlett	<.001		<.001	
	Variance explained	72.114		69.411	

5: Motorway exit, rural road	PW	**.824**	.169	**.839**	.162
	NASA TLX	**.862**	-.056	**.931**	.020
	STRESS	**.846**	-.095	**.907**	.090
	SCR	.361	**.733**	.164	**.709**
	HR	-.320	**.734**	-.008	**.799**
	KMO	.711		.706	
	Bartlett	<.001		<.001	
	Variance explained	69.729		71,914	
6: Village, rural road to outskirts of town	PW	**.831**	.042	**.836**	.207
	NASA TLX	**.879**	-.018	**.922**	.057
	STRESS	**.844**	-.153	**.908**	.031
	SCR	.171	**.833**	.274	**.708**
	HR	-.263	**.714**	-.069	**.834**
	KMO	.684		.706	
	Bartlett	<.001		<.001	
	Variance explained	70.019		73.955	
7 City	PW	**.834**	.155	**.834**	.191
	NASA TLX	**.934**	-.004	**.934**	-.020
	STRESS	**.906**	.117	**.916**	.043
	SCR	.207	**.735**	.029	**.818**
	HR	-.039	**.837**	.093	**.790**
	KMO	.700		.690	
	Bartlett	<.001		<.001	
	Variance explained	74.224		74.986	

As Table 2 demonstrates, the stable two-dimensional factor structure (PR and CR) remains remarkably consistent regardless of the specific route segment (e.g., rural road vs. city traffic) or the environment (real-world vs. simulator). Consequently, the structural stability of the composite indicators is fully supported (H1).

Load-Response Relationship (Testing H2)

We tested H2 by correlating the composite indicators with Vehicle Operation (VO), the measure of situational load. As expected, VO correlated negatively with both composite indicators (indicating that lower capability leads to higher stress). In the simulator drive, CR correlated strongly with VO ($r = -.646$, $p < .001$) and PR correlated significantly with VO ($r = -.142$, $p < .01$). In the real drive, CR ($r = -.402$, $p < .001$) and PR ($r = -.246$, $p < .01$) also showed significant correlations. Comparing this to single measures highlights the value of composite indicators: In the simulator condition, while SCR correlated with VO ($r = -.125$), HR alone did not ($r = -.064$, ns). By aggregating them into PR, the composite indicator captured the shared stress variance and yielded a significant result where single measures failed. Consequently, H2 is supported.

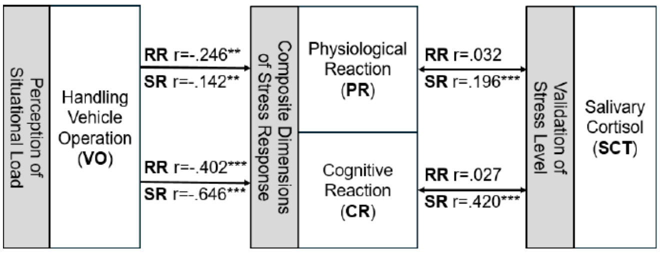
Biological Validation (Testing H3)

Finally, we tested the biological validity of these composite indicators (H3). In the simulator condition (which successfully induced stress), both composite indicators showed highly significant correlations with cortisol levels (see Figure 1):

- **PR and SCT**: r = .196, p < .001; **CR and SCT**: r = .420, p < .001.

In the real drive, correlations were non-significant (PR r = .032; CR r = .027). This is likely due to the "floor effect" of stress in the real drive; without sufficient variance in stress activation, covariance with cortisol cannot be established. However, the significant results in the simulator drive provide support for the validity of the composite indicators. The fact that CR correlated more strongly with cortisol (r=.420) than PR (r=.196) suggests that cognitive appraisal is a potent proxy for biological stress outcomes in this context, aligning with theoretical predictions by [8].

Figure 1. Final research model and key statistical results (RR/SR = real ride/simulated ride)



Discussion and Conclusion

Theoretical Contributions

This study contributes to the NeuroIS field by demonstrating that composite indicators offer a more reliable basis for interpretation than single metrics. First, we showed that Physiological Reaction (PR), combining HR and SCR, remains stable even when individual signals fluctuate due to situational artifacts. This addresses the "noisy data" problem in applied NeuroIS research, where movement or sensor detachments can ruin single-channel data [4]. Second, the strong correlation of Cognitive Reaction (CR) with cortisol suggests that simple, aggregated self-reports can effectively proxy biological stress when physiological measurement is not feasible. Interestingly, cortisol correlates more strongly with CR (r = .420) than with PR (r = .196). This conceptually aligns with the two primary stress pathways: While PR captures the rapid, often unspecific arousal of the Symptahtetic Adrenal Medullary (SAM) axis, cortisol is regulated by the Hypothalamus-Pituitary-Adrenal (HPA) axis. The HPA axis is highly sensitive to cognitive appraisal and activates primarily when a situation is perceived as demanding or overwhelming [15]. Since CR explicitly measures this subjective evaluation, its strong correlation with cortisol is theoretically sound. This supports the "measurement pluralism" perspective [23], showing that subjective and objective measures are not opposing forces but complementary parts of a holistic stress assessment.

Limitations, Future Research and Conclusion

Several limitations must be noted. First, the real drive induced minimal stress, limiting the validation potential in that condition. Future studies should experimentally manipulate demand levels more aggressively in real-world settings to test if the correlations hold. Second, the order of conditions was fixed (real then sim), which might interact with cortisol's diurnal cycle. Third, while driving is a key application area, future work should validate these composite indicators in other IS contexts, such as human-AI interaction or mobile system usage, to ensure the PR/CR structure generalizes beyond the automotive domain [24]. Fourth, while our results demonstrate the value of the Cognitive Reaction (CR) factor, it is critical to acknowledge the inherent methodological weaknesses of self-reports in NeuroIS research. As highlighted by Tams et al. (2014) [3] and vom Brocke et al. (2020) [25], participants often struggle to accurately recall and reflect upon their true affective responses. Furthermore, repeatedly querying users after each stimulus can introduce severe priming biases, altering their subsequent behavior. Finally, we must consider the 'contrary perspective' where self-reports are simply not feasible. In many naturalistic IS use-cases, such as continuous high-speed driving, interrupting the user with a questionnaire would destroy the task context or endanger the user. In such scenarios, the continuous, non-intrusive measurement provided by the Physiological Reaction (PR) composite indicator becomes not just a methodological alternative, but an absolute necessity.

Single physiological and cognitive indicators are often insufficient for reliable user experience measurement in NeuroIS. This study demonstrates that composite indicators, strategically aggregating GSR, HR, and self-report measures, provide greater stability, validity, and interpretability. We validated that the constructs of Physiological Reaction and Cognitive Reaction remain stable across environments and correlate with biological ground truth (cortisol) under load. We therefore recommend the use of PR and CR composite indicators as a standard methodological basis for analyzing activation and demand in future NeuroIS user tests.

References

1. Riedl R, Léger P-M. Fundamentals of NeuroIS. Berlin, Heidelberg: Springer Berlin Heidelberg; 2016.
2. Dimoka A, Banker RD, Benbasat I, Davis FD, Dennis AR, Gefen D, et al. On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for Neurois1. MISQ. 2012;36:679–702. doi:10.2307/41703475.
3. Tams S, Hill K, Guinea A, Thatcher J, Grover V. NeuroIS—Alternative or Complement to Existing Methods? Illustrating the Holistic Effects of Neuroscience and Self-Reported Data in the Context of Technostress Research. JAIS. 2014;15:723–53. doi:10.17705/1jais.00374.
4. Lohani M, Payne BR, Strayer DL. A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. Front Hum Neurosci. 2019;13:57. doi:10.3389/fnhum.2019.00057.

5. Gradl S, Wirth M, Richer R, Rohleder N, Eskofier BM. An Overview of the Feasibility of Permanent, Real-Time, Unobtrusive Stress Measurement with Current Wearables. In: Mayora O, Forti S, Meyer J, Mamykina L, editors. PervasiveHealth'19: The 13th International Conference on Pervasive Computing Technologies for Healthcare; 20 05 2019 23 05 2019; Trento Italy. New York, NY, USA: ACM; 2019. p. 360–365. doi:10.1145/3329189.3329233.
6. Andreassi JL. Psychophysiology: Psychology Press; 2010.
7. Riedl R. On the biology of technostress. SIGMIS Database. 2012;44:18–55. doi:10.1145/2436239.2436242.
8. Cohen S, Gianaros PJ, Manuck SB. A Stage Model of Stress and Disease. Perspect Psychol Sci. 2016;11:456–63. doi:10.1177/1745691616646305.
9. Boucsein W. Electrodermal Activity. Boston, MA: Springer US; 2012.
10. Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. IEEE Trans Inf Technol Biomed. 2010;14:410–7. doi:10.1109/TITB.2009.2036164.
11. Folkman S. Stress: Appraisal and Coping. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer New York; 2013. p. 1913–1915. doi:10.1007/978-1-4419-1005-9_215.
12. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52:281–302. doi:10.1037/h0040957.
13. Rushton JP, Brainerd CJ, Pressley M. Behavioral development and construct validity: The principle of aggregation. Psychol Bull. 1983;94:18–38. doi:10.1037/0033-2909.94.1.18.
14. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: recent developments and applications. Psychoneuroendocrinology. 1994;19:313–33. doi:10.1016/0306-4530(94)90013-2.
15. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull. 2004;130:355–91. doi:10.1037/0033-2909.130.3.355.
16. Czaban M, Himmels C. Investigating simulator validity by using physiological and cognitive stress indicators. Transportation Research Part F: Traffic Psychology and Behaviour. 2025;114:831–51. doi:10.1016/j.trf.2025.07.006.
17. Czaban M, Sultanow E, Chircu A, Czarnecki C, Riedl J, Wengler S. User Interaction with Digital Twins for Driving: How Comparable Are Simulated and Real Trajectories? Systems. 2026;14:144. doi:10.3390/systems14020144.
18. Wynne RA, Beanland V, Salmon PM. Systematic review of driving simulator validation studies. Safety Science. 2019;117:138–51. doi:10.1016/j.ssci.2019.04.004.
19. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Human Mental Workload: Elsevier; 1988. p. 139–183. doi:10.1016/S0166-4115(08)62386-9.
20. Helton WS, Näswall K. Short Stress State Questionnaire. European Journal of Psychological Assessment. 2015;31:20–30. doi:10.1027/1015-5759/a000200.
21. Lewis JR. Measuring User Experience With 3, 5, 7, or 11 Points : Does It Matter? Hum Factors. 2021;63:999–1011. doi:10.1177/0018720819881312.

22. Healey JA, Picard RW. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. IEEE Trans. Intell. Transport. Syst. 2005;6:156–66. doi:10.1109/TITS.2005.848368.
23. Fischer T, Riedl R. Technostress Research: A Nurturing Ground for Measurement Pluralism? CAIS. 2017;40:375–401. doi:10.17705/1CAIS.04017.
24. Chen L, Zhao Y, Ye P, Zhang J, Zou J. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Systems with Applications. 2017;85:279–91. doi:10.1016/j.eswa.2017.01.040.
25. vom Brocke J, Hevner A, Léger PM, Walla P, Riedl R. Advancing a NeuroIS research agenda with four areas of societal contributions. European Journal of Information Systems. 2020;29:9–24. doi:10.1080/0960085X.2019.1708218.

Recovery of Physiological Videoconferencing Fatigue: An Expectancy-Violations Perspective

Johannes Bernard[1,*] , Christoph Weinert[1], and Tim Weitzel[1]

[1] Otto-Friedrich University, Chair of Information Systems and Services, Bamberg, Germany
{johannes.bernard,christoph.weinert,tim.weitzel}@uni-bamberg.de

Abstract. Physiological videoconference fatigue (PVCF) has emerged as a significant threat to employee well-being. While current research focusses on the drivers of PVCF, this study integrates expectancy-violations theory to examine how prior expectations regarding videoconferences influence subsequent physiological recovery to homeostasis. We conducted a 2x3 between-subject laboratory experiment and measured PVCF through skin conductance within a simulated videoconference environment. Preliminary results indicate that higher amplitudes in physiological arousals do not necessarily lead to longer recovery times and that violated expectations lead to longer recovery times than confirmed expectations.

Keywords: videoconference fatigue · expectancy-violations theory · arousal · skin conductance · recovery

Introduction

Recent statistics from 2024 show that employees spent almost one-third of their work time videoconferencing [1]. Besides its positive effects, studies have also shown that the excessive use of videoconferencing tools (e.g., MS Teams) is associated with negative, stress-related consequences – known as physiological videoconference fatigue (PVCF) [2, 3]. PVCF is defined as the resulting state of physical and mental exhaustion [4]. While videoconferencing leads to activation of the sympathetic nervous system (SNS) [4] to prepare the body for action [5], resulting in biological user responses (e.g., physiological arousals), PVCF is associated with increased activity of the parasympathetic nervous system (PNS) [3, 6]. Prior research shows that the state of fatigue is reflected by decreased physiological responses [3]. PVCF is among others caused by information overload [7] such as self-awareness, interaction with multiple faces and multitasking, and demands a higher cognitive effort than face-to-face communication, which eventually leads to fatigue-related consequences [3, 4].

However, the processing of information depends on individuals' expectations [8], and its assessment depends on individuals' experiences [9]. How users respond to information overload during videoconferences depend on their expectations before entering the conference. Based on expectancy-violations theory (EVT; [10]), we assume that violations and confirmations regarding the expectations towards an upcoming videoconference lead to different physiological responses. Expectations can either be met (confirmation) or unmet (violation). In the PVCF context, prior experiences lead to the development of expectations [11] that get violated, leading to information overload [12] and activation of the SNS [10, 13]. Moreover, past research on PVCF has focused on the occurrence of the response [2, 3]. Yet, it is also important to understand how quickly users recover from fatigue. The human body responds by trying to recover from these physiological arousals and return to a stable and constant condition, which is called homeostasis where SNS and PNS are in an equilibrium [14, 15]. The duration of recovery varies [16], with shorter recovery times implying being able to achieve higher cognitive performance earlier again [17]. Research has shown that expectation violations demand more cognitive effort [10, 18], which we assume to result in longer PVCF recovery. Therefore, we differentiate between two phases, the initial SNS activation caused by expectation violations, and subsequently the activation of the PNS caused by PVCF. We assume that PVCF recovery differs depending on violations and confirmations *regarding information overload* [10]. Despite research of PVCF drivers, current literature lacks an integration of how expectation confirmations/violations influence PVCF recovery. Therefore, the research question is:

How do expectation violations regarding information overload while videoconferencing result in different physiological VCF recovery?

Theoretical Background and Hypotheses Development

Expectancy-Violations Theory

EVT explains how individuals evaluate the outcomes from social interactions and states that individuals develop expectations toward a social interaction which either get confirmed or violated, leading to different outcomes [10].

Beyond influencing immediate responses, EVT suggests that expectancy violations trigger ongoing cognitive evaluation processes. These processes may persist beyond the initial interaction and thereby influence the temporal dynamics of physiological recovery, rather than solely affecting the magnitude of initial responses.

Physiological Videoconference Fatigue and Recovery

Videoconference fatigue (VCF) has emerged as a phenomenon in recent research and is defined as a state of physical and mental exhaustion caused by excessive use of videoconferencing tools [4]. Most research about VCF focused on the pure existence of VCF and its root causes [4, 19] and it was not until Riedl et al. (2023) explored VCF from a neurophysiological perspective that this view expanded.[2] PVCF extends self-reported feelings of VCF by also emphasizing physiological measurable effects such as electrodermal activity, heart rate, and eye-related metrics [2, 3]. Recent work has found further support for these mechanisms through EEG and autonomic nervous system evidence, highlighting the complex biological toll of VCF [6].

[2] Additionally, both terms are synonyms and often used interchangeably [4].

Just like VCF, PVCF is caused by information overload [4, 7, 19]. Because videoconferencing lacks the intuitive cues of face-to-face communication, the brain must increase cognitive effort to compensate for its unique features, resulting in PVCF [4]. Information overload in videoconferences is caused by three antecedents: self-awareness, unnatural interaction with multiple faces, and multitasking opportunities. *Self-awareness* arises from the fact that videoconferencing tools provide the participants with the opportunity to constantly see themselves like they are looking at a mirror [19], leading to increased self-awareness opportunities and the secretion of stress hormones [4, 19, 20]. Unnatural interaction with *multiple faces* arises from the typical arrangement of the other participants' faces in videoconferences. While direct eye contact is not constantly used in face-to-face communication [19], the interface of videoconferencing tools make it seem like the participant is being stared at constantly by multiple other participants [4], and the feeling of being stared at is known to trigger physiological arousals [21, 22]. The opportunity to *multitask* while participating in a videoconference is another root cause of information overload [4]. Studies show that this simultaneous involvement in multiple activities and rapid switching between multiple windows is associated with increased perceived stress and physiological responses [23, 24]. It is expected that participants have expectations regarding the three factors of information overload when joining a videoconference, which can either get confirmed or violated.

We differentiate between two distinct phases reflecting the dynamic interplay of the autonomic nervous system [5]: (i) the initial SNS activation caused by an expectation-violation regarding the three factors of information overload during videoconferences, and (ii) subsequently the activation of the PNS caused by PVCF during the videoconference. The first phase comprises physiological responses to videoconferencing. Videoconferencing environments can elicit acute physiological arousal due to cognitive and social demands, which is associated with activation of the SNS. When being confronted with expectation violations the SNS gets activated, leading to physiological arousal, indicated by biomarkers such as skin conductance (SC) elevation [15, 25].

The second phase is characterized by prolonged videoconferencing leading to fatigue-related state, which have been associated with increased PNS activity, reflecting regulatory efforts to restore homeostasis [3, 6]. One mechanism to reach homeostasis is physiological recovery. Physiological recovery means returning to homeostasis after expectation violation induced SNS activation through subsequent PNS activation. In the context of electrodermal activity it is reflected by a return of the physiological arousal to its baseline level [16]. The present study aims to connect the two phases by examining how the first phase of expectation violations, which cause activation of the SNS and lead to physiological arousals impacts the second phase of physiological recovery and PVCF, which are associated with PNS activity [10, 13]. Hence, this study does not directly measure PVCF as a fatigue state but focuses on physiological recovery following videoconference-induced arousal, which may represent a key mechanism underlying the emergence of PVCF. Shorter recovery times indicate more efficient autonomic regulation, whereas prolonged recovery may indicate sustained physiological engagement and delayed return to homeostasis.

Hypotheses development

In the following section, we will theoretically develop our hypotheses regarding the effects of expectation violations when videoconferencing and their subsequent physiological recovery.

Because expectation violations require continued cognitive processing to resolve discrepancies [10], they are expected to prolong physiological activation beyond the initial response phase, thus leading to longer recovery. An overview of our hypothesized effects in our research model is provided by *Figure 1*.

While increased self-awareness is causing information overload while videoconferencing [4], EVT suggests that the intensity of the resulting physiological response is depending on prior expectations [10]. When a participants' expectation regarding self-awareness is confirmed, the cognitive system acts in a standard pattern within an expected framework [26], resulting in minimized cognitive effort [10, 27] compared to violations [18]. Such an expectation violation reflects phase one, the activation of the SNS; followed by the second phase, which is associated with PNS activity and eventually leads to return to homeostasis. Seeing oneself in a 'mirror' while videoconferencing requires the brain to work harder, which is assumed to results in physiological arousal [4, 19]. During violations, the brain needs to work harder to adjust to the unexpected situation [18], which is why we assume longer recovery. Hence, we hypothesize:

H1: Violated expectations regarding self-awareness while videoconferencing lead to longer recovery than confirmed expectations regarding self-awareness while videoconferencing.

The human brain has been hardwired for face-to-face communication in natural group settings. In videoconferences, it feels like multiple people are staring at you at the same time, which might result in physiological arousal [3, 4], providing a baseline of potential physiological arousal. When participants of a videoconference expect this, their expectations are confirmed, suggesting that their brain can process automatically with lower cognitive effort [10, 27], leading to activation of the SNS and physiological arousals [16], from which the body recovers over time to return to homeostasis [15, 25]. Since expectation violations are associated with higher, continued cognitive effort than confirmations, it is assumed that expectation violations lead to longer recovery. Hence, we hypothesize:

H2: Violated expectations regarding the interaction with multiple faces while videoconferencing lead to longer recovery than confirmed expectations regarding the interaction with multiple faces while videoconferencing.

Similarly, engaging in multiple tasks while videoconferencing leads to stress and physiological arousal [4, 24]. Confirmed expectations regarding multitasking (e.g. participants have to multitask while videoconferencing, but they expected it) still lead to physiological arousals caused by SNS activation from which the body tries to recover [15]. Violated expectations regarding multitasking lead to increased cognitive effort, as the brain needs more resources to cope with this violation [10, 18]. During this process, the brain must work harder to return to homeostasis [10, 15, 18], assumably leading to prolonged recovery. Hence, we hypothesize:

H3: Violated expectations regarding multitasking while videoconferencing lead to longer recovery than confirmed expectations regarding multitasking while videoconferencing.

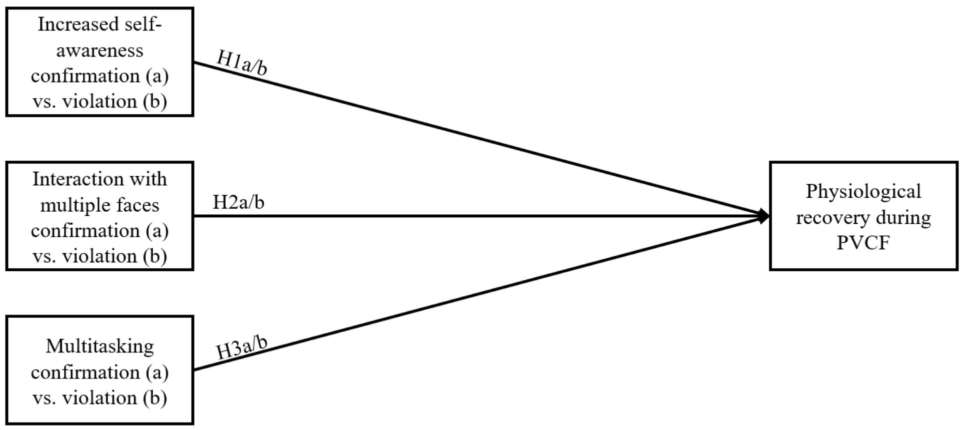


Fig. 7. Research Model.

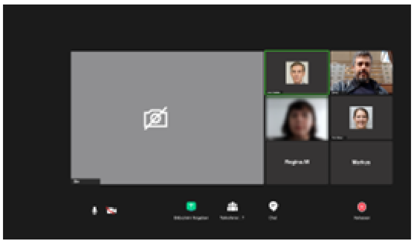
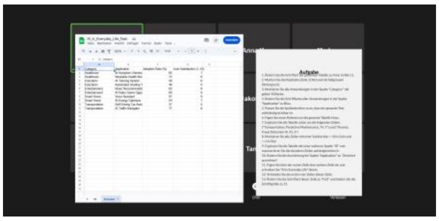
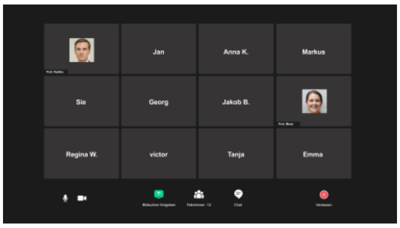
Research Methodology and Experimental Procedure

We developed a 2x3 between-subject laboratory experiment. The first factor was expectation, consisting of confirmed and violated [10]. The second factor was information overload [4], consisting of self-awareness, interaction with multiple faces, and multitasking. The design results in six experimental conditions: (1) confirmed self-awareness, (2) violated self-awareness, (3) confirmed interaction with multiple faces, (4) violated interaction with multiple faces, (5) confirmed multitasking, and (6) violated multitasking.

Manipulations: The self-awareness manipulation is achieved by showing the participants their appearance on the majority of the simulated videoconference screen [4]. The interaction with multiple faces manipulation is achieved by showing the participants 20 faces on the videoconference screen to give the impression of being stared at concurrently by 20 people [4]. The multitasking manipulation is simulated by giving the participants a task while participating in the videoconference. *Table 1* provides an overview of the manipulations within the simulated videoconference environment. For the simulated videoconference, we created a 12min long conversation between two people. Additionally, to ensure ecological validity, we introduced participants to imagine they were participating in a work-related videoconference about “AI in the daily life”, where two experts were talking about this topic. Lastly, during the post-experimental stage, we asked the participants questions about the content of the videoconference to measure performance [28].

Manipulation check: We recorded participants’ head movement while participating in the videoconference to ensure they were looking at the manipulation. Participants also had to report if they were aware whether they saw their own camera image, how many other participants had their cameras on, and if applicable, how they divided their attention between the task and the videoconference.

Tab. 1. Overview of Manipulations.

Information overload antecedent	
Self-awareness (self-view on the left side)	**Multiple Faces**
	
Multitasking	**Control**
	

Measurement: To measure the participants physiological responses, the participants were wearing a mobile device (MentalBioScreen K3) that recorded their SC values in microsiemens (μS). SC measurement is well suited to capture autonomic nervous system activity [29, 30], thus applicable to capture both physiological arousal caused by expectation violations through SNS activation and subsequent PVCF and physiological recovery which are associated with PNS activity [3, 6].

Experimental procedure: The procedure was split into three parts (see *Figure 2*). During the pre-experimental stage, the participants were asked to self-report their perceived VCF using the zoom fatigue scale [31]. We also asked the participants to report their expectations regarding the antecedents of information overload, so they could either get confirmed or violated. We captured the SC baseline level of all participants during a rest phase in this stage [16]. During the experimental stage, the participants joined the simulated videoconference. This stage captures the two phases. First, expectation violations and information overload are assumed to cause acute physiological arousal, reflected in activation of the SNS and increases in SC response amplitude. Second, following this initial activation, PVCF and physiological recovery processes associated with the PNS are assumed to facilitate the return toward baseline levels. To measure participants' recovery from physiological arousal, we applied the 50% and 63% approaches, which refer to the time span until the electrodermal level recovers 50% or 63% of the peak amplitude after a stimulus [16]. These approaches are commonly used, as participants are often unlikely to reach the pre-arousal again due to baseline shifts [16]. To create the simulated videoconference and make the experience as realistic as possible – thus further strengthening ecological validity [30], we used Labvanced to replicate the visual appearance of MS Teams. Additionally, the manipulations were contextually embedded [30].

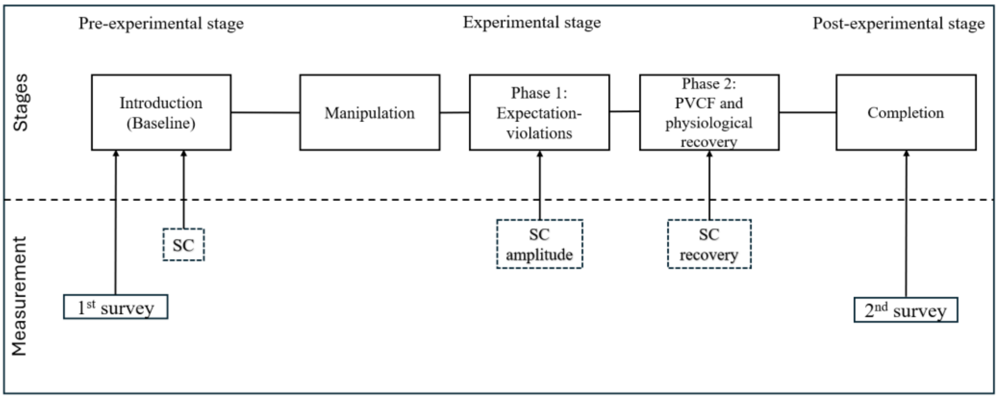


Fig. 8. Experimental Procedure.

First Results

To test the experiment, we conducted a pre-test with seven participants. The pre-test results shown in *Table 2* indicate that higher SCR amplitudes do not necessarily translate to longer recovery times. Violated expectations regarding interaction with multiple faces led to higher SCR amplitudes and the participants did not recover during the experiment, indicating that violated expectations regarding the information overload antecedents do indeed lead to longer recovery times, indicating support for H2.

Tab. 2. Overview of Preliminary Results.

(N = 7)	**Confirmation**			**Violation**		
Information overload antecedent	**SCR amplitude in µS**	**50% SCR rec. in s**	**63% SCR rec. in s**	**SCR amplitude in µS**	**50% SCR rec. in s**	**63% SCR rec. in s**
Self-awareness	155.38	21	48.5	-	-	-
Multiple Faces	38.05	156	192	110.15	NR	NR
Multitasking	-	-	-	40.83	84.5	107
Control*	66.05	15	17			
Note: SCR amplitude: vertical height of a response relative to the baseline; 50% (63%) SCR rec.: time required for the SCR curve to recover 50% (63%) of its amplitude [16]; NR: not reached within experimental window; * = no stressor						

Expected Contributions and Next Steps

The results of this study are expected to contribute to literature in the following ways. First, we expect to contribute to the field of NeuroIS [29, 32] by incorporating a physiological recovery perspective on PVCF. While existing NeuroIS research in the context of PVCF focuses on autonomic nervous system activity [2, 3, 6], we extend this literature by incorporating subsequent physiological recovery, shifting the focus from mere stress induction to the body's ability to return to homeostasis. This enables differentiation between increased SNS activation and impaired recovery caused by PVCF, allowing for further differentiation of activation and deactivation processes of the autonomic nervous system [5].

Second, we expect to advance current understanding of PVCF information overload antecedents [4] by introducing an individual, situational dimension. Through incorporating EVT [10], we demonstrate that PVCF is regulated by user expectations and their confirmation or violation. Specifically, first results showed that expectation violations regarding these antecedents lead to longer recovery than confirmations, suggesting that the psychological, situational dimension of videoconferences influences PVCF and PVCF recovery. This reveals that the three information overload antecedents are only part of the reason why PVCF happens, with prior expectations regarding the videoconference and psychological processes as the other part.

Building on these preliminary results, we will recruit a larger sample to validate the research model, as well as include additional neurophysiological measures such as heart rate and heart rate variability. In doing so, we will be able to gain even richer insights into autonomic nervous system activity and physiological recovery [29].

References

1. Erkic, A.: Meeting Statistics You Should Know (2024). (2024)
2. Pagel, L.V., Kojic, T., Schmitt, V., Wardah, W., Harnisch, P.L., Möller, S., Spang, R.P.: Signals of Strain: Using Biosignals to Assess Fatigue in Video Conferencing. In: Harris, D. and Li, W.-C. (eds) Engineering Psychology and Cognitive Ergonomics. pp. 118–133. Springer Nature Switzerland, Cham (2025)
3. Riedl, R., Kostoglou, K., Wriessnegger, S.C., Müller-Putz, G.R.: Videoconference fatigue from a neurophysiological perspective: experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). Sci Rep. 13, 18371 (2023). https://doi.org/10.1038/s41598-023-45374-y
4. Riedl, R.: On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. Electron Markets. 32, 153–177 (2022). https://doi.org/10.1007/s12525-021-00501-3
5. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS: Information Systems and the Brain. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
6. Whelan, E., Riedl, R., Salo, M., Pirkkalainen, H.: Virtual meetings and your brain: Four ways to refresh. MIT Sloan Management Review. (2025)

7. Kock, N.: The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution. Organization Science. 15, 327–348 (2004). https://doi.org/10.1287/orsc.1040.0071
8. Rossel, P., Peyrin, C., Kauffmann, L.: Subjective perception of objects depends on the interaction between the validity of context-based expectations and signal reliability. Vision Research. 206, 108191 (2023). https://doi.org/10.1016/j.visres.2023.108191
9. Arnold, M., Goldschmitt, M., Rigotti, T.: Dealing with information overload: a comprehensive review. Frontiers in psychology. 14, 1122200 (2023)
10. Burgoon, J.K.: Expectancy Violations Theory. In: The International Encyclopedia of Interpersonal Communication. pp. 1–9. John Wiley & Sons, Ltd (2015)
11. Churchill Jr, G.A., Surprenant, C.: An investigation into the determinants of customer satisfaction. Journal of marketing research. 19, 491–504 (1982)
12. Bock, G.-W., Mahmood, M., Sharma, S., Kang, Y.J.: The Impact of Information Overload and Contribution Overload on Continued Usage of Electronic Knowledge Repositories. Journal of Organizational Computing and Electronic Commerce. 20, 257–278 (2010). https://doi.org/10.1080/10919392.2010.494530
13. Le Poire, B.A., Burgoon, J.K.: Usefulness of differentiating arousal responses within communication theories: Orienting response or defensive arousal within nonverbal theories of expectancy violation. Communication Monographs. 63, 208–230 (1996). https://doi.org/10.1080/03637759609376390
14. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress from a Neurobiological Perspective: System Breakdown Increases the Stress Hormone Cortisol in Computer Users. Bus Inf Syst Eng. 4, 61–69 (2012). https://doi.org/10.1007/s12599-012-0207-7
15. Riedl, R.: On the biology of technostress: literature review and research agenda. SIGMIS Database. 44, 18–55 (2012). https://doi.org/10.1145/2436239.2436242
16. Boucsein, W.: Electrodermal Activity. Springer Science & Business Media (2012)
17. Marko, M., Riečanský, I.: Sympathetic arousal, but not disturbed executive functioning, mediates the impairment of cognitive flexibility under stress. Cognition. 174, 94–102 (2018). https://doi.org/10.1016/j.cognition.2018.02.004
18. Mendes, W.B., Blascovich, J., Hunter, S.B., Lickel, B., Jost, J.T.: Threatened by the unexpected: Physiological responses during social interactions with expectancy-violating partners. Journal of Personality and Social Psychology. 92, 698–716 (2007). https://doi.org/10.1037/0022-3514.92.4.698
19. Bailenson, J.N.: Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. Technology, Mind, and Behavior. 2, (2021). https://doi.org/10.1037/tmb0000030
20. Slavich, G.M., Way, B.M., Eisenberger, N.I., Taylor, S.E.: Neural sensitivity to social rejection is associated with inflammatory responses to social stress. Proc. Natl. Acad. Sci. U.S.A. 107, 14817–14822 (2010). https://doi.org/10.1073/pnas.1009164107
21. Senju, A., Johnson, M.H.: The eye contact effect: mechanisms and development. Trends in cognitive sciences. 13, 127–134 (2009)

22. Takac, M., Collett, J., Blom, K.J., Conduit, R., Rehm, I., De Foe, A.: Public speaking anxiety decreases within repeated virtual reality training sessions. PloS one. 14, e0216288 (2019)
23. Reinecke, L., Aufenanger, S., Beutel, M.E., Dreier, M., Quiring, O., Stark, B., Wölfling, K., Müller, K.W.: 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, 90–115 (2017). https://doi.org/10.1080/15213269.2015.1121832
24. Wetherell, M.A., Carter, K.: The Multitasking Framework: The Effects of Increasing Workload on Acute Psychobiological Stress Reactivity. Stress and Health. 30, 103–109 (2014). https://doi.org/10.1002/smi.2496
25. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress aus einer neurobiologischen Perspektive: Systemabsturz führt bei Computerbenutzern zu einem Anstieg des Stresshormons Kortisol. Wirtschaftsinf. 54, 59–68 (2012). https://doi.org/10.1007/s11576-012-0314-6
26. Burgoon, J.K.: Interpersonal Expectations, Expectancy Violations, and Emotional Communication. Journal of Language and Social Psychology. 12, 30–48 (1993). https://doi.org/10.1177/0261927X93121003
27. Samuel, J., Kashyap, R., Samuel, Y., Pelaez, A.: Adaptive cognitive fit: Artificial intelligence augmented management of information facets and representations. International Journal of Information Management. 65, 102505 (2022). https://doi.org/10.1016/j.ijinfomgt.2022.102505
28. Madsen, J., Parra, L.C.: Attentional state, not trait, predicts test performance in video-based learning. iScience. 28, 113622 (2025). https://doi.org/10.1016/j.isci.2025.113622
29. 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.: On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for Neurois. MIS quarterly. 36, 679–702 (2012)
30. Balapour, A., Riedl, R.: Ecological Validity in NeuroIS Research: Theory, Evidence, and a Roadmap for Future Studies. Journal of the Association for Information Systems. 26, 9–65 (2025). https://doi.org/10.17705/1jais.00904
31. Fauville, G., Luo, M., Queiroz, A.C.M., Bailenson, J.N., Hancock, J.: Zoom Exhaustion & Fatigue Scale. Computers in Human Behavior Reports. 4, 100119 (2021). https://doi.org/10.1016/j.chbr.2021.100119
32. Riedl, R., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P.: On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. Communications of the Association for Information Systems. 27, 15 (2010)

Human vs. Automated Feedback: A Study on the Dissociation Between Physiological Activation and Cognitive Self-Focus

Sophie Bisson[1], Mario Passalacqua[2], Laurent Joblot[3], Florian Magnani[4], Sylvain Sénécal[5], Constantinos K. Coursaris[1], Robert Pellerin[6], and Pierre-Majorique Léger[1]

[1] Department of Information Technologies, Tech3Lab, HEC Montréal, Canada
{sophie.bisson, constantinos.coursaris, pierre-majorique.leger}@hec.ca
[2] Department of Psychology, Université du Québec à Montréal, Canada
passalacqua.mario@uqam.ca
[3] LISPEN / HESAM University / UBFC, Arts et Métiers Institute of Technology, Chalon-sur-Saône, France
laurent.joblot@ensam.eu
[4] iaelyon, Université Jean Moulin Lyon 3, France
florian.magnani@univ-lyon3.fr
[5] Department of Marketing, Tech3Lab, HEC Montréal, Canada
sylvain.senecal@hec.ca
[6] Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Canada
robert.pellerin@polymtl.ca

Abstract. This study examines how feedback delivery mode shapes operators' physiological activation and self-focus during a repetitive industrial task. Using a multi-method NeuroIS approach (N = 31), participants completed a time-constrained disassembly task while receiving periodic performance feedback delivered either verbally by a researcher or visually via a digital interface. Results reveal a distinct dissociation: automated feedback significantly increased physiological activation (heart rate), while human feedback triggered higher cognitive self-focus (self-reflection). Self-efficacy emerged as a significant covariate, indicating meaningful individual differences consistent with transactional stress theory. These findings suggest that automated interfaces exert a persistent monitoring pressure, whereas human presence introduces social-evaluative processing. This research provides design implications for balancing physiological activation and cognitive load in smart factory environments.

Keywords: Physiological Activation, Cognitive Self-Focus, Industry 5.0, Automated Feedback, Human Feedback, Self-Efficacy.

Introduction

The digital transformation of industrial workplaces has reshaped performance monitoring through the integration of artificial intelligence (AI). This shift is central to Industry 5.0 (I5.0), a paradigm that moves beyond technical efficiency to prioritize worker well-being, empowerment, and human-centric research [1]. AI-driven feedback systems are increasingly embedded in performance infrastructures [2–5], intended to augment rather than replace workers [6, 7]. While automation improves routine performance [8, 9], it can create "out-of-the-loop" problems during failure [10, 11]. Despite the established technical efficacy [2, 3, 5, 12], the psychological consequences of feedback source (human vs. automated) remain underexplored in NeuroIS [6, 7, 13].

Performance feedback regulates behavior by signaling discrepancies between current and desired performance, shaping effort allocation and strategy adjustment [12, 14–17]. However, feedback source is not neutral; it influences interpretation, trust, and internalization [2–5, 12]. The same message may elicit distinct psychological responses depending on whether it originates from a human supervisor or an automated interface.

Drawing from Self-Determination Theory (SDT), which posits that optimal functioning depends on autonomy, competence, and relatedness [4, 18, 19] and the Trilogy of Mind framework distinguishing affective, motivational, and cognitive domains [20, 21], we argue that feedback source activates distinct stress pathways. We define physiological activation as autonomic arousal indexed by heart rate, representing the affective/arousal pathway [21–23]. Within the NeuroIS field, HR is recognized as a foundational tool for capturing objective neurophysiological data during system interaction [24]. Recent neuroscientific evidence suggests that cardiac sympathetic-vagal activity plays a leading and causal role in initiating the emotional response, where ascending modulations from heartbeat activity precede and sustain cortical responses to arousal [25]. Furthermore, HR has been validated as a sensitive metric for tracking moment-by-moment changes in affective processing during continuous task execution [26].

Human feedback conveys social presence and relatedness [2, 5, 12], supporting psychological needs and goal internalization. Yet its social-evaluative nature can heighten self-consciousness and cognitive self-focus [3, 12, 27, 28], redirecting attention toward self-monitoring rather than task execution [29, 30]. Automated feedback, often perceived as objective and socially distant [5], may reduce evaluative self-consciousness but can function as a continuous, impersonal monitoring force [2–5]. Lacking social nuance or pauses, it may sustain physiological activation even without interpersonal judgment [3, 31].

Responses likely vary across individuals. Consistent with the Transactional Model of Stress (TMS), stress reflects appraisals of demands relative to coping resources [32, 33]. Self-efficacy, belief in one’s capability to execute task demands, shapes these appraisals. High self-efficacy may render feedback informational; low self-efficacy may make it threatening [4, 18, 19, 34]. Physiologically, these appraisals modulate autonomic reactivity [35]; while a 'threat' appraisal triggers sharp sympathetic activation and heightened heart rate [25], high self-efficacy serves as a physiological buffer that maintains autonomic stability by framing the feedback as a manageable challenge [34–36].

Using a multi-method NeuroIS approach, we distinguish physiological activation (heart rate) from cognitive self-focus (self-report) to advance understanding of operator states in I5.0. We hypothesize: **(H1)** participants receiving human feedback will report higher cognitive self-focus compared to those receiving automated feedback; **(H2)** automated feedback will be associated with higher physiological activation compared to human feedback; **(H3)** self-efficacy will significantly contribute to these responses, serving as a buffer against physiological activation and cognitive self-focus.

Methods

This section details the experimental design and multi-method measures, following Passalacqua et al. [37].

Participants

Thirty-four participants were recruited; after artifact screening, the final sample included 31 (12 F; age M = 28.38, SD = 11.78). Inclusion required normal or corrected vision and no prior task experience. Ethical approval was obtained (certificate #2023-5427); all participants provided written consent.

Experimental Design and Task

A between-subjects design compared human (n = 15) and automated (n = 16) feedback. Participants performed a repetitive snowshoe disassembly task under a challenging time goal. Feedback (ahead/behind pace) was delivered every two minutes: verbally by a researcher (human condition) or via a standardized interface (automated condition). The between-subjects design was selected to prevent carryover and contamination effects. Specifically, the repetitive nature of the task could introduce learning effects in a within-subjects setup [38], while prior research shows that the presence of a human observer induces self-directed attention and performance regulation processes [39], making it likely that such cognitive effects would influence subsequent performance across conditions.

Instruments and Measures

Continuous ECG was recorded at 256 Hz using the Hexoskin Smart Garments (Carré Technologies Inc., Montreal, Canada). A 1-minute standing resting baseline was recorded prior to the task to allow for physiological stabilization. Data were processed using Hexoskin's proprietary algorithms, which include automated R-peak detection and signal quality indicators. Segments with low signal quality were excluded from analysis to reduce the influence of noise and movement artifacts on cardiovascular metrics. Mean heart rate (bpm) indexed sympathetic activation [22]. Baseline self-efficacy was measured with the General Self-Efficacy Short Scale – French Version (ASKU-F, 3-item French adaptation;[40]). Post-task workload and stress were assessed using NASA Task Load Index (NASA-TLX; [41]) and a reduced 12-item Short Stress State Questionnaire (SSSQ; [35]), selecting the two highest-loading items per factor to preserve structure while reducing burden.

Statistical Analysis

Hypotheses were tested via ANCOVA in SAS (Type III sums of squares), with baseline self-efficacy (ASKU) included as a covariate to control for individual differences in physiological activation. Given the directional nature of our hypotheses and an analytical sample of $N = 31$, one-tailed tests ($\alpha = .05$) were evaluated. Because no formal a priori power analysis was conducted, a sensitivity analysis (G*Power) was performed to estimate the minimum detectable effect size.

For the primary ANCOVA effect (power = .80, 1 covariate), the design was sensitive to effects of at least Cohen's $f = 0.52$ (equivalent to $\eta^2 = 0.21$), confirming that the study was adequately sensitive to detect the large stress-pathway dissociations observed in the final model ($R^2 = 0.317$).

Results

Physiological Activation

ANCOVA on mean heart rate revealed a significant feedback source effect, $F(1, 26) = 3.49$, $p = .030$ (one-tailed). Supporting H2, automated feedback produced higher activation ($M = 97.74$, $SE = 4.31$) than human feedback ($M = 86.37$, $SE = 4.46$). Self-efficacy significantly predicted heart rate, $F(1, 26) = 5.43$, $p = .028$; higher self-efficacy was associated with lower heart rate ($\beta = -9.03$, $p = .028$), where higher levels of self-efficacy were associated with lower mean heart rates. This supports H3, indicating a physiological buffering effect.

Cognitive Self-focus

Due to low internal consistency ($\alpha = 0.59$), SSSQ self-reflection items were analyzed separately. A significant effect emerged for "I am reflecting about myself," $F(1, 26) = 4.42$, $p = .023$ (one-tailed), supporting H1. "I am self-conscious" showed a marginal effect, $F(1, 27) = 2.49$, $p = .063$ (one-tailed). No NASA-TLX differences emerged (ps $> .05$). Self-efficacy did not significantly predict cognitive self-focus ($p > .05$).

Discussion

Findings indicate a dissociation between physiological and cognitive stress pathways as a function of feedback source, consistent with the Trilogy of Mind [21].

Automated Feedback and Physiological Activation (H2)

Support for H2 suggests that automated monitoring is associated with increased physiological activation. This pattern may reflect the continuous and system-paced nature of digital interfaces, which can be perceived as persistent evaluation [3, 4]. Unlike human interaction, which includes social pauses and contextual cues, automated feedback delivers consistent performance signals that may sustain generalized activation. In Industry 5.0 (I5.0) contexts, the limited social framing of automated systems may contribute to prolonged physiological engagement [2–5], aligning with neuroergonomics research documenting elevated load in human–robot environments [6, 7].

Human Feedback and Cognitive Self-Focus (H1)

Results for H1 suggest that human feedback may shift individuals toward a more cognitive-evaluative state. Drawing on SDT [19], human interaction can support relatedness and warmth [2, 5, 12]. At the same time, social presence may increase awareness of being evaluated. This may redirect attentional resources toward self-monitoring or impression management [29, 30]. Alternatively, social facilitation research indicates that the mere presence of an observer can heighten public self-consciousness even in the absence of explicit threat [27, 28]. Thus, the human element in I5.0 appears to introduce additional cognitive processing demands, distinct from purely physiological activation.

The Buffering Role of Self-Efficacy (H3)

Higher self-efficacy was significantly associated with lower mean heart rates, supporting its role as a physiological buffer. This finding is consistent with the TMS [36], suggesting that confident operators may appraise feedback as manageable rather than threatening. However, self-efficacy did not significantly reduce cognitive self-reflection. This pattern indicates that physiological stability does not necessarily eliminate social-evaluative processing [42]. Self-efficacy therefore appears to function as a situational cognitive resource that moderates physiological activation, distinct from SDT's broader competence need.

Practical Implications for Industry 5.0

These findings suggest that feedback systems in I5.0 should account for distinct stress pathways. Automated systems may benefit from supportive framing that reduces perceptions of continuous monitoring [2, 4, 5], and positioning AI as a decision aid rather than a rigid selector may support autonomy. Human supervisors, conversely, may need to be attentive to unintended social-evaluative pressure and adopt development-oriented feedback approaches consistent with SDT [15, 18, 19]. Sustainable performance in smart factories may depend on balancing algorithmic precision with mechanisms that support operator self-efficacy.

Limitations and Future Research

The modest sample size ($N = 31$) may limit sensitivity to smaller effects. Additionally, the use of university students may constrain generalizability, as experienced factory workers may differ in adaptation to monitoring systems. Future longitudinal research is needed to assess whether this dissociation persists over extended work periods or long-term AI integration. Further work could also examine alternative feedback framings, including more socially attuned automated systems and emerging technologies such as digital twins or augmented reality [3, 5, 6, 12].

References

1. European Commission.: Industry 5.0: towards a sustainable, human centric and resilient European industry. Publications Office, LU (2021).
2. Pei, J., Wang, H., Peng, Q., Liu, S.: Saving face: Leveraging artificial intelligence-based negative feedback to enhance employee job performance. Hum. Resour. Manage. 63, 775–790 (2024). https://doi.org/10.1002/hrm.22226.
3. Thuillard, S., Adams, M., Jelmini, G., Schmutz, S., Sonderegger, A., Sauer, J.: When humans and computers induce social stress through negative feedback: Effects on performance and subjective state. Comput. Hum. Behav. 133, 107270 (2022). https://doi.org/10.1016/j.chb.2022.107270.
4. Edwards, M.R., Zubielevitch, E., Okimoto, T., Parker, S., Anseel, F.: Managerial control or feedback provision: How perceptions of algorithmic HR systems shape employee motivation, behavior, and well-being. Hum. Resour. Manage. 63, 691–710 (2024). https://doi.org/10.1002/hrm.22218.
5. Hein, I., Cecil, J., Lermer, E.: Acceptance and motivational effect of AI-driven feedback in the workplace: an experimental study with direct replication. Front. Organ. Psychol. 2, 1468907 (2024). https://doi.org/10.3389/forgp.2024.1468907.
6. Roy, R.N., Drougard, N., Gateau, T., Dehais, F., Chanel, C.P.C.: How Can Physiological Computing Benefit Human-Robot Interaction? Robotics. 9, 100 (2020). https://doi.org/10.3390/robotics9040100.
7. Zakeri, Z., Arif, A., Omurtag, A., Breedon, P., Khalid, A.: Multimodal Assessment of Cognitive Workload Using Neural, Subjective and Behavioural Measures in Smart Factory Settings. Sensors. 23, 8926 (2023). https://doi.org/10.3390/s23218926.
8. Calzavara, M., Faccio, M., Granata, I., Trevisani, A.: Achieving productivity and operator well-being: a dynamic task allocation strategy for collaborative assembly systems in Industry 5.0. Int. J. Adv. Manuf. Technol. (2024). https://doi.org/10.1007/s00170-024-14302-3.
9. Daniel Ajiga, Patrick Azuka Okeleke, Samuel Olaoluwa Folorunsho, Chinedu Ezeigweneme: The role of software automation in improving industrial operations and efficiency. Int. J. Eng. Res. Updat. 7, 022–035 (2024). https://doi.org/10.53430/ijeru.2024.7.1.0031.
10. Bainbridge, L.: Ironies of Automation. Automatica. 19, 775–779 (1983).
11. Endsley, M.R., Kiris, E.O.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Hum. Factors J. Hum. Factors Ergon. Soc. 37, 381–394 (1995). https://doi.org/10.1518/001872095779064555.
12. Marx, C., Könczöl, C., Altmanninger, A., Kubicek, B.: The Critical Robot: Impact of Performance Feedback on Intrinsic Motivation, Self-Esteem and Psychophysiology in Human–Robot Interaction. Int. J. Soc. Robot. 16, 1751–1764 (2024). https://doi.org/10.1007/s12369-024-01147-9.

13. Nuamah, J.K., Mantooth, W., Karthikeyan, R., Mehta, R.K., Ryu, S.C.: Neural Efficiency of Human–Robotic Feedback Modalities Under Stress Differs With Gender. Front. Hum. Neurosci. 13, 287 (2019). https://doi.org/10.3389/fnhum.2019.00287.
14. Robison, M.K., Unsworth, N., Brewer, G.A.: Examining the effects of goal-setting, feedback, and incentives on sustained attention.
15. Krenn, B., Würth, S., Hergovich, A.: The Impact of Feedback on Goal Setting and Task Performance: Testing the Feedback Intervention Theory. Swiss J. Psychol. 72, 79–89 (2013). https://doi.org/10.1024/1421-0185/a000101.
16. Locke, E.A., Latham, G.P.: The development of goal setting theory: A half century retrospective. Motiv. Sci. 5, 93–105 (2019). https://doi.org/10.1037/mot0000127.
17. Murthy, U.S., Schafer, B.A.: The Effects of Relative Performance Information and Framed Information Systems Feedback on Performance in a Production Task. J. Inf. Syst. 25, 159–184 (2011). https://doi.org/10.2308/jis.2011.25.1.159.
18. Beenen, G., Pichler, S., Levy, P.E.: Self-Determined Feedback Seeking: The Role of Perceived Supervisor Autonomy Support. Hum. Resour. Manage. 56, 555–569 (2017). https://doi.org/10.1002/hrm.21787.
19. Laguerre, R.A., Barnes-Farrell, J.L.: Bringing Self-Determination Theory to the Forefront: Examining How Human Resource Practices Motivate Employees of All Ages to Succeed. J. Bus. Psychol. 40, 1–37 (2025). https://doi.org/10.1007/s10869-024-09951-w.
20. Helton, W.S., Näswall, K.: Short Stress State Questionnaire: Factor Structure and State Change Assessment. Eur. J. Psychol. Assess. 31, 20–30 (2015). https://doi.org/10.1027/1015-5759/a000200.
21. Hilgard, E.R.: The trilogy of mind: Cognition, affection, and conation. J. Hist. Behav. Sci. 16, 107–117 (1980). https://doi.org/10.1002/1520-6696(198004)16:2%253C107::AID-JHBS2300160202%253E3.0.CO;2-Y.
22. Cherif, N.H., Mezghani, N., Gaudreault, N., Ouakrim, Y., Mouzoune, I., Boulay, P.: Physiological Data Validation of the Hexoskin Smart Textile: In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies. pp. 150–156. SCITEPRESS - Science and Technology Publications, Funchal, Madeira, Portugal (2018). https://doi.org/10.5220/0006588001500156.
23. Passalacqua, M., Pellerin, R., Yahia, E., Magnani, F., Rosin, F., Joblot, L., Léger, P.-M.: Practice With Less AI Makes Perfect: Partially Automated AI During Training Leads to Better Worker Motivation, Engagement, and Skill Acquisition." International Journal of Human–Computer Interaction, March 3, 2024, 1–21. https://doi.org/10.1080/10447318.2024.2319914. Int. J. Human–Computer Interact. 1–21 (2024). https://doi.org/10.1080/10447318.2024.2319914.
24. 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.: ON THE USE OF NEUROPHYSIOLOGICAL TOOLS IN IS RESEARCH: DEVELOPING A RESEARCH AGENDA FOR NEUROIS.

25. Candia-Rivera, D., Catrambone, V., Thayer, J.F., Gentili, C., Valenza, G.: Cardiac sympathetic-vagal activity initiates a functional brain–body response to emotional arousal. Proc. Natl. Acad. Sci. 119, e2119599119 (2022). https://doi.org/10.1073/pnas.2119599119.
26. Shehu, H.A., Oxner, M., Browne, W.N., Eisenbarth, H.: Prediction of moment-by-moment heart rate and skin conductance changes in the context of varying emotional arousal. Psychophysiology. 60, e14303 (2023). https://doi.org/10.1111/psyp.14303.
27. Belletier, C., Normand, A., Huguet, P.: Social-Facilitation-and-Impairment Effects: From Motivation to Cognition and the Social Brain. Curr. Dir. Psychol. Sci. 28, 260–265 (2019). https://doi.org/10.1177/0963721419829699.
28. Fenigstein, A.: Self-Consciousness, Self-Attention, and Social Interaction.
29. Muraven, M.: Self-Focused Attention and the Self-Regulation of Attention: Implications for Personality and Pathology. J. Soc. Clin. Psychol. 24, 382–400 (2005). https://doi.org/10.1521/jscp.24.3.382.65615.
30. Spurr, J.M., Stopa, L.: Self-focused attention in social phobia and social anxiety. Clin. Psychol. Rev. 22, 947–975 (2002). https://doi.org/10.1016/S0272-7358(02)00107-1.
31. Taskasaplidis, G., Fotiadis, D.A., Bamidis, P.D.: Review of Stress Detection Methods Using Wearable Sensors. IEEE Access. 12, 38219–38246 (2024). https://doi.org/10.1109/ACCESS.2024.3373010.
32. Winstone, N.E., Hepper, E.G., Nash, R.A.: Individual differences in self-reported use of assessment feedback: the mediating role of feedback beliefs. Educ. Psychol. 41, 844–862 (2021). https://doi.org/10.1080/01443410.2019.1693510.
33. Klingbeil, A., Grützner, C., Schreck, P.: Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. Comput. Hum. Behav. 160, 108352 (2024). https://doi.org/10.1016/j.chb.2024.108352.
34. Wiczorek, R., Meyer, J.: Effects of Trust, Self-Confidence, and Feedback on the Use of Decision Automation. Front. Psychol. 10, 519 (2019). https://doi.org/10.3389/fpsyg.2019.00519.
35. Gellman, M.D., Turner, J.R. eds: Encyclopedia of Behavioral Medicine. Springer New York, New York, NY (2013). https://doi.org/10.1007/978-1-4419-1005-9.
36. Obbarius, N., Fischer, F., Liegl, G., Obbarius, A., Rose, M.: A Modified Version of the Transactional Stress Concept According to Lazarus and Folkman Was Confirmed in a Psychosomatic Inpatient Sample. Front. Psychol. 12, 584333 (2021). https://doi.org/10.3389/fpsyg.2021.584333.
37. Passalacqua, M., Pellerin, R., Magnani, F., Joblot, L., Yahia, E., Rosin, F., Léger, P.-M.: L’origine de l’objectif est-elle importante? Effets motivationnels d’objectifs autodéfinis en production. In: Proceedings of CIGI Qualita MOSIM 2023. , Trois-Rivières, QC, Canada (2023). https://doi.org/10.60662/f8tk-ms61.
38. Walsh, M.M., Krusmark, M.A., Jastrembski, T., Hansen, D.A., Honn, K.A., Gunzelmann, G.: Enhancing learning and retention through the distribution of practice repetitions across multiple sessions. Mem. Cognit. 51, 455–472 (2023). https://doi.org/10.3758/s13421-022-01361-8.

39. Carver, C.S., Scheier, M.F.: The self-attention-induced feedback loop and social facilitation. J. Exp. Soc. Psychol. 17, 545–568 (1981). https://doi.org/10.1016/0022-1031(81)90039-1.
40. Décieux, J.P., Sischka, P.E., Schumacher, A., Willems, H.: FASKU. ZIS - Collect. Items Scales Soc. Sci. (2020). https://doi.org/10.6102/ZIS290_EXZ.
41. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, 10.1016/S0166-4115(08)62386-9, (1988).
42. Gellman, M.D., Turner, J.R. eds: Encyclopedia of Behavioral Medicine. Springer New York, New York, NY (2013). https://doi.org/10.1007/978-1-4419-1005-9.

Gastric Myoelectrical Dynamics During Experimentally Induced Affective and Cognitive States

Meghana Vadakkedath Dharmapalan[1✉], Alexander Maedche[1], and Benjamin Scheibehenne[1]

[1] Karlsruhe Institute of Technology (KIT), Kaiserstr. 89, 76133 Karlsruhe, Germany
{meghana.dharmapalan,alexander.maedche,benjamin.scheibehenne}@kit.edu

Abstract. This study investigated whether experimentally induced affective and cognitive states can be detected from gastric myoelectrical activity, a physiological signal that has received comparatively limited attention in state-detection research despite its important role within the gut–brain axis. In a laboratory-controlled within-subject design (N = 77), participants underwent fear, disgust, relaxation, and cognitive load (2-back) conditions while electrogastrogram (EGG) signals were recorded, and frequency-based gastric features were extracted and compared using repeated-measures ANOVAs. Most gastric features did not vary significantly across conditions; however, fear was associated with increased total gastric power and dynamic dominant power without shifts in dominant frequency, indicating selective modulation of gastric amplitude during emotional arousal.

Keywords: Electrogastrography · Psychological states · Gut–brain axis

Introduction

Psychological states are known to influence autonomic regulation of the body's visceral organs. Our study examined whether psychological states can be detected from gastric myoelectrical activity. There has been a long tradition of investigating biomarkers of psychological states in various physiological signals such as electrocardiogram, electrodermal activity, and electroencephalogram data [1-3]. Within the NeuroIS field, these signals have been adopted as tools to capture cognitive and affective processes during human-computer interaction [4, 5]. In contrast, the relationship between gastric electrophysiology and these processes has received comparatively limited attention.

Autonomic modulation of gastric activity occurs via the gut-brain axis (GBA) [6], a bidirectional communication pathway between the central nervous system and the enteric nervous system. Psychological states shift the autonomic balance, directly influencing gastrointestinal motility and gastric myoelectrical patterns [7 - 9]. Psychological states thus may manifest as measurable variations in gastric myoelectrical activity.

In line with this assumption, recent brain-gut coupling studies found phase-locked interactions between gastric rhythms and cortical networks [10 - 13], with stronger coupling associated with worse mental health signatures (anxiety, depression, stress, and wellbeing)[13]. A similar relationship is reported between gastric disorders and depression, anxiety, and other mental health indicators. [14,15]. However, most research focuses on chronic conditions or long-term associations [13 - 15] rather than experimentally induced psychological states.

Previous experimental studies reported reduced gastric activity during emotion induction [16,17]. However, these studies mainly compared broad positive vs. negative emotions and rarely included non-emotional cognitive states, and findings have been inconsistent [18,19]. Further work is therefore needed to examine how experimentally induced affective and cognitive states influence gastric activity. This is particularly relevant as understanding how distinct affective and cognitive states manifest in physiological signals is a central concern of NeuroIS research [5].

Towards this goal, our present study investigated the following research question: To what extent do specific affective (fear, disgust, relaxation) and cognitive (load) states differentially modulate gastric myoelectrical parameters? We induced affective and cognitive states that are known to engage autonomic processes. Affective states of fear, disgust, and relaxation were chosen because fear is associated with increased sympathetic activation and altered parasympathetic activity [20], disgust is strongly linked to visceral and gastric sensations [21], and relaxation served as a neutral baseline condition. Cognitive load was chosen as it engages executive control processes with minimal affective arousal [22]. We measured electrogastrogram (EGG) data, and electrocardiogram (ECG) data served as a manipulation check as it is an established proxy for ANS modulation and for psychological states [3]. To assess state-dependent differences, specific gastric signal features were extracted and statistically compared across conditions.

Method

To investigate whether psychological states can be detected in gastric activity we conducted a laboratory-controlled, within-subject study where the participants were exposed to four conditions designed to induce distinct psychological states (fear, disgust, relaxation, and cognitive load). The conditions, fear, disgust, and cognitive load were counterbalanced across participants, with relaxation interleaved serving as the baseline. This structure ensured that every induction was preceded and followed by a baseline phase to allow for physiological recovery. Emotional states were induced using validated five-minute video clips. Cognitive load was induced using the 2-back working memory task [23] where the participants had to memorize a sequence of letters and respond by pressing the ‘J’ key when the current letter had appeared two trials ago, and ‘F’ key when it was a new letter. Following each video, the participants rated the intensity of four experienced emotions (fear, disgust, relaxation, surprise) during the video. After the 2-back task, the participants rated the task difficulty, fatigue, and frustration. Each condition lasted 5 minutes, and the full experimental session lasted approximately 40 minutes. We chose the five-minute induction, to prevent attenuation of the induced emotion which occurs in longer durations, while still keeping the total session under 40 minutes to avoid participant fatigue.

Participants

Ninety-nine participants were recruited from a university student pool (70 male, 29 female), with most aged 18 – 24 age bracket. Participants at the time of the study were not taking medication for neurological, psychiatric, gastrointestinal, or cardiovascular conditions. Inclusion criteria required a body mass index (BMI) of 18–26 kg/m² and a minimum 2-hour fast before the session. The study had received ethical approval from the Institutional Review Board.

Procedure

Upon arrival at the lab, the participants were informed about the study, after which they provided written informed consent. The sensors were placed on the participants. Skin preparation was performed using alcohol swabs prior to electrode attachment. Participants were seated in an office chair with a reclined backrest (135°) and instructed to remain as still as possible. After entering age and gender, participants began the experimental task.

EGG and ECG data were recorded continuously for the whole duration of the experiment. Accelerometer (ACC) sensors were also used to check movement contamination.

Physiological Recording

Gastric myoelectrical activity was acquired using a Biosignal Plux electrogastrogram (EGG) device [24]. The EGG sensors used a uni-channel bipolar configuration, consisting of two active electrode channels and a reference electrode. Sensors were placed non-invasively on the upper abdominal region over the stomach following Rizzo et al. configuration [25]. Cardiac activity was recorded using a Biosignal Plux electrocardiogram (ECG) sensor [26]. The ECG sensors also had a uni-channel, bipolar electrode configuration and were placed on the participants according to Lead I Einthoven configuration [27]. Movement was recorded using Biosignal Plux tri-axial sensing ACC sensors [28] that capture acceleration along three axes. Two ACC sensors were used; one placed on the abdomen to capture trunk movement, and another placed on the right inner forearm to monitor hand movement. All physiological signals were acquired using the Biosignal Plux 8-channel hub [29] and OpenSignals (r)evolution software [30]. The sampling rate was set to 300 Hz to accommodate both EGG and ECG signal acquisition.

Data Processing and Analysis

Of the 99 participants recruited, 16 were excluded due to non-systematic software issues involving physiological signal acquisition and data-saving errors. An additional six datasets were discarded due to a software failure that resulted in missing behavioral data. Consequently, a total of 77 complete behavioral and physiological datasets were included in the final analysis.

Self-report ratings of emotions experienced during the conditions were analysed to verify the effectiveness of the psychological manipulations. To ensure that physiological effects were not attributable to movement artifacts, signal quality was assessed prior to feature extraction.

EGG Processing. EGG signals were preprocessed following Wolpert's recommendations for EGG signal analysis [31]. Gastric waves are slow rhythmic oscillations occurring at approximately 2-4 cycles per minute (cpm) in healthy adults. Signals were band-pass filtered between 0.5 and 9 cpm to attenuate movement artifacts and high-frequency noise. Gastric signals can be characterised in time or frequency domain. As gastric signals are low-amplitude and susceptible to motion artifacts, time features are difficult to interpret, hence we focused on frequency-domain features. To extract the features, spectral decomposition was performed using Fast Fourier Transform (FFT). Features were computed for each full condition segment (5 minutes) and additionally using a 60-second sliding window to capture dynamic changes across time within each condition. The following frequency-domain features were extracted for each condition: dominant frequency, total spectral power, and the percentage of spectral power within the normogastric waves (2-4 cpm), bradygastric (1-2 cpm), and tachygastric (4-9 cpm) bands. Dominant frequency reflects the frequency at which maximal spectral power occurs within the gastric range, while total spectral power represents the overall magnitude of rhythmic gastric activity. Band-specific power percentages quantify the relative distribution of gastric activity across physiologically defined frequency ranges. Additionally, for each window, the dominant frequency was identified and its corresponding spectral power calculated. Instability coefficients were calculated for both dominant frequency and dominant power as the coefficient of variation across windows. The proportion of windows classified within bradygastric, normogastric, and tachygastric bands was computed to quantify time spent in each rhythm range within each condition. These features were selected based on commonly reported features in prior EGG studies [32].

ECG Processing. ECG signals were processed using NeuroKit2 [33]. R-peak detection was performed using the Hamilton (2002) algorithm [34], and signal quality was assessed using the averageQRS and Zhao (2018) methods [35]. HRV features extracted per condition included mean heart rate, SDNN, RMSSD, pNN50, and spectral power in the VLF, LF and HF bands, normalised LF and HF, total power, with LF/HF ratio calculated as an additional index of autonomic balance.

Statistical Testing. To test whether experimentally induced psychological states were associated with differences in gastric and cardiac features, separate repeated-measures ANOVAs were conducted for each physiological feature, with phase (fear, disgust, relaxation, cognitive load) as the within-subject factor. Sphericity was assessed using Mauchly's test and was violated for 24 of 28 features, therefore Greenhouse-Geisser corrected degrees of freedom and p-values are reported throughout. Normality was assessed using the Shapiro-Wilk test, which indicated violations across features and conditions. Given the sample size of N = 77, repeated-measures ANOVA was considered sufficiently robust to normality violations [36, 37]. When significant omnibus effects were observed, post-hoc pairwise comparisons were performed. False Discovery Rate (FDR) correction using the Benjamini–Hochberg procedure was applied to control for multiple comparisons. Data was processed in Python, data handling performed using pandas [38] and NumPy [39]. Signal processing was conducted using SciPy [40] and NeuroKit2.

Results

Participants rated the intensity of four emotions after each condition on a scale of 1 to 4. Ratings were generally consistent with the intended conditions. Fear (M = 3.16, SD = 1.38) and disgust (M = 3.23, SD = 1.41) ratings were highest in their respective conditions, and relaxation was highest during baseline (M = 3.95, SD = 1.29). The n-back task was rated as moderately difficult (M = 2.82, SD = 0.81), tiring (M = 2.05, SD = 0.91), and frustrating (M = 2.55, SD = 0.93).

A series of repeated-measures ANOVAs examined physiological variation across the four phases (Baseline, Fear, Disgust, and cognitive load) for each gastric and cardiac feature (29 features in total). Most gastric features did not show significant phase effects. However, a significant phase effect was seen for total gastric power ($F(1.87, 141.82) = 6.77$, $p_{FDR} = .006$, $\eta 2 = .08$) and mean dynamic dominant power ($F(1.87, 142.33) = 6.8$, $p_{FDR} = .006$, $\eta 2 = .08$).

Following FDR correction for multiple comparisons, post-hoc comparisons revealed that fear was associated with increased gastric power and dynamic dominant power (both $p_{FDR} = .009$) relative to baseline, without significant shifts in dominant frequency ($p_{FDR} > .05$)

Significant phase effects were observed for mean heart rate ($F(2.7, 206.34) = 4.01$, $p_{FDR} = .02$, $\eta 2 = .05$) and multiple HRV features, SDNN ($F(2.8, 172.4) = 4.44$, $p_{FDR} = .02$, $\eta 2 = .05$), RMSSD ($F(2.4, 170.26) = 4.72$, $p_{FDR} = .02$, $\eta 2 = .05$), pNN50 ($F(2.9, 220.61) = 7.67$, $p_{FDR} < .001$, $\eta 2 = .09$), TP ($F(2.85, 216.7) = 7.98$, $p_{FDR} < .001$, $\eta 2 = .09$), VLF ($F(3, 228) = 7.42$, $p_{FDR} < .001$, $\eta 2 = .08$), LF ($F(3, 228) = 13$, $p_{FDR} < .001$, $\eta 2 = .14$), LF normalised ($F(3, 228) = 11.29$, $p_{FDR} < .001$, $\eta 2 = .12$), HF ($F(2.8, 211.2) = 4.83$, $p_{FDR} = .01$, $\eta 2 = .05$), HF normalised ($F(3, 228) = 6.12$, $p_{FDR} = .002$, $\eta 2 = .07$), and LF/HF ratio ($F(3, 228) = 7.3$, $p_{FDR} < .001$, $\eta 2 = .08$).

Heart rate increased during cognitive load (2-back) relative to baseline ($p_{FDR} = .009$). Cognitive load was further associated with reductions in total power ($p_{FDR} < .001$), LF ($p_{FDR} < .001$), normalised LF ($p_{FDR} < .001$), HF ($p_{FDR} = .02$), and LF/HF ratio ($p_{FDR} < .001$). Fear condition was associated with reductions in VLF ($p_{FDR} = < .001$), LF ($p_{FDR} = .02$), LF/HF ratio ($p_{FDR} = .01$), alongside increases in normalized HF ($p_{FDR} = .002$). Additionally, disgust was associated with a decrease in VLF ($p_{FDR} = .01$) and LF/HF ratio (p_{FDR} = .04), and increase in pNN50 (p_{FDR} = .002) and normalised HF (p_{FDR} = .001).

Discussion

The present study investigated whether gastric physiological activity varies across distinct psychological states. Gastric measures showed selective modulation during fear, though the majority of features were not found to be significant, suggesting limited sensitivity across the tested conditions. Fear was associated with increased total gastric power and mean dominant power relative to baseline, without significant shifts in dominant frequency or band distribution. This pattern may suggest modulation of gastric wave amplitude rather than their frequency composition. The findings therefore tentatively suggest that emotional arousal can influence gastric activity without necessarily producing frequency changes or overt rhythm disruption. This trend is contrary to previous findings and classical theories [16,17] that propose gastric suppression during sympathetic activation. It might also be possible that we are seeing the initial stage of gastric modulation that leads to gastric suppression. Future work should examine whether longer or more intense emotional inductions produce different patterns of gastric response.

In contrast to gastric activity, cardiac measures demonstrated broader sensitivity to the psychological phase. Cognitive load was associated with increased heart rate and reductions in multiple HRV indices, including LF, normalised LF, HF, total power, and LF/HF ratio, reflecting reduced overall variability during sustained cognitive demand. Emotional conditions produced more selective spectral changes, particularly reductions in LF/HF ratio alongside increases in normalized HF power. These findings suggest that cardiac indices track both cognitive and affective demands, reflecting generalized autonomic adjustments across psychological contexts.

While the increase in heart rate, indicating general arousal, is consistent with previous research [3], the observed increases in HRV indices typically associated with parasympathetic activity during the psychological states diverge from conventional findings. These unexpected findings warrant cautious interpretation and further investigation. While HRV indices are known to be sensitive to factors such as caffeine intake and physical fitness [41], these were not strictly controlled in the present study and may have contributed to the observed variability in cardiac responses.

Together, these results suggest differential sensitivity of cardiac and gastric systems to psychological state. Cardiac measures showed variation across both cognitive and emotional conditions, whereas gastric modulation appeared more specific to fear-related emotional processing. Notably, cognitive load did not significantly alter gastric power despite variation in cardiac signals. This pattern might suggest that gastric myoelectrical dynamics may be more closely linked to affective arousal than to non-emotional cognitive effort, though this interpretation remains preliminary and requires replication and further investigation.

The present findings reflect modulation under brief, laboratory-controlled inductions of psychological states. The five-minute induction periods were designed to capture transient state-dependent changes; however, longer or more sustained manipulations may reveal slower gastric dynamics not observable within this timeframe. Variability in signal quality and individual differences in gastric sensitivity may also contribute to heterogeneity in gastric responses. Additionally, gender-level analysis was not conducted as significant gender-based differences in gastric activity have not been previously established for the conditions tested. Furthermore, lifestyle factors including caffeine intake, smoking, and physical fitness were not strictly controlled, which may have introduced variability in cardiac measures. Future research could employ extended induction protocols with stricter controls for lifestyle factors and a more gender-balanced sample. Systematically examining individual differences in gut–brain coupling would further clarify the conditions under which gastric myoelectrical activity is sensitive to psychological modulation.

References

1. Lin, W., Li, C.: Review of studies on emotion recognition and judgment based on physiological signals. Applied Sciences. 13, 4, 2573 (2023)
2. Sequeira, H., Hot, P., Silvert, L., Delplanque, S.: Electrical autonomic correlates of emotion. Int. J. Psychophysiol. 71(1), 50–56 (2009)
3. Bulagang, A.F., Weng, N.G., Mountstephens, J., Teo, J.: A review of recent approaches for emotion classification using electrocardiography and electrodermography signals. Inform. Med. Unlocked 20, 100363 (2020)
4. 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.
5. Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS: Information systems and the brain. Springer.
6. Carabotti, M., Scirocco, A., Maselli, M.A., Severi, C.: The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Ann. Gastroenterol. 28(2), 203–209 (2015)
7. Vianna, E.P.M., Tranel, D.: Gastric myoelectrical activity as an index of emotional arousal. Int. J. Psychophysiol. 61(1), 70–76 (2006)
8. Vianna, E.P.M., Naqvi, N., Bechara, A., Tranel, D.: Does vivid emotional imagery depend on body signals? Int. J. Psychophysiol. 72(1), 46–50 (2009)
9. Muth, E.R., Koch, K.L., Stern, R.M., Thayer, J.F.: Effect of autonomic nervous system manipulations on gastric myoelectrical activity and emotional responses in healthy human subjects. Psychosom. Med. 61(3), 297–303 (1999)
10. Rebollo, I., Wolpert, N., Tallon-Baudry, C.: Brain–stomach coupling: Anatomy, functions, and future avenues of research. Current Opinion in Biomedical Engineering. 18, 100270 (2021)
11. Rebollo, I., Devauchelle, A.D., Béranger, B., Tallon-Baudry, C.: Stomach-brain synchrony reveals a novel, delayed-connectivity resting-state network in humans. eLife 7, e33321 (2018)
12. Azzalini, D., Rebollo, I., Tallon-Baudry, C.: Visceral signals shape brain dynamics and cognition. Trends in cognitive Science. 23(6), 488–509 (2019)
13. Banellis, L., Rebollo, I., Nikolova, N., Allen, M.: Stomach–brain coupling indexes a dimensional signature of mental health. Nature Mental Health. 3(8), 899–908 (2025)
14. Koloski, N.A., Jones, M., Kalantar, J., Weltman, M., Zaguirre, J., Talley, N.J.: The brain–gut pathway in functional gastrointestinal disorders is bidirectional: a 12-year prospective population-based study. Gut. 61(9), 1284–1290 (2012)

15. Walker, E.A., Katon, W.J., Jemelka, R.P., Roy-Byrne, P.P.: Comorbidity of gastrointestinal complaints, depression, and anxiety in the Epidemiologic Catchment Area (ECA) Study. The American journal of medicine. 92(1), S26–S30 (1992)
16. Baldaro, B., Battacchi, M.W., Codispoti, M., Tuozzi, G., Trombini, G., Bolzani, R., Palomba, D.: Modifications of electrogastrographic activity during the viewing of brief film sequences. Perceptual and motor skills. 82(3_suppl), 1243–1250 (1996)
17. Zhou, R., Luo, Y., Hu, S.: Effects of viewing pleasant and unpleasant photographs on gastric motility indexed by electrogastrographic (EGG) activities. Perceptual motor skills. 99(3), 785–789 (2004)
18. Baldaro, B., Mazzetti, M., Codispoti, M., Tuozzi, G., Bolzani, R., Trombini, G.: Autonomic reactivity during viewing of an unpleasant film. Perceptual motor skills. 93(3), 797–805 (2001)
19. Valentina, Z., Mario, V., Juliane, C., Vincent, N., Sami, A., Morwena, L.: The gastric myoelectrical activity response to emotional stimuli. In: 2018 IX International Seminar of Biomedical Engineering (SIB), pp. 1–5. IEEE (2018)
20. Jang, E.H., Byun, S., Park, M.S., Sohn, J.H.: Predicting individuals' experienced fear from multimodal physiological responses to a fear-inducing stimulus. Advances in cognitive psychology. 16(4), 291 (2020)
21. Ottaviani, C., Mancini, F., Petrocchi, N., Medea, B., Couyoumdjian, A.: Autonomic correlates of physical and moral disgust. Int. J. Psychophysiol. 89(1), 57–62 (2013)
22. Hansen, A.L., Johnsen, B.H., Thayer, J.F.: Vagal influence on working memory and attention. Int. J. Psychophysiol. 48(3), 263–274 (2003)
23. Kirchner, W.K.: Age differences in short-term retention of rapidly changing information. J. Exp. Psychol. 55(4), 352 (1958)
24. Electrogastrography (EGG) Sensor, https://www.pluxbiosignals.com/products/electrogastrography-egg?srsltid=AfmBOooqVtrj9fvFp3Hft-WMyaMdvbtjHi_v3jDuhW6redaQO0P1NL35c
25. Riezzo, G., Russo, F., Indrio, F.: Electrogastrography in adults and children: the strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity. BioMed research international. 282757 (2013).
26. Electrocardiography (ECG) Sensor, https://www.pluxbiosignals.com/products/electrocardiography-ecg-sensor-1?srsltid=AfmBOooKHO4OuXnoZ4c8e6Iw1kpw-ud-DuAIVxZgE4u0eogrrk2oXnQHD
27. Einthoven, W.: The different forms of the human electrocardiogram and their signification. The Lancet. 179(4622), 853–861 (1912)
28. Accelerometer (ACC), https://www.pluxbiosignals.com/products/accelerometer-acc?srsltid=AfmBOoogo41gZJn4I35Un11sY5sr1zVKLBE9sUIghX6wH42CJjdQQTaf
29. 8-Channel biosignalsplux Kit, https://www.pluxbiosignals.com/products/8-channel-biosignals-kit?srsltid=AfmBOopl6gn2yTcK-BIvEfyAps84KWVm82bu0ZfoeL4A8W6dZtvxLNkoH
30. OpenSignals (r)evolution (Download), https://support.pluxbiosignals.com/knowledge-base/introducing-opensignals-revolution/
31. Wolpert, N., Rebollo, I., Tallon-Baudry, C.: Electrogastrography for psychophysiological research: Practical considerations, analysis pipeline, and normative data in a large sample. Psychophysiology. 57(9), e13599 (2020)
32. Kim, N., Seo, W., Kim, S., Park, S.M.: Electrogastrogram: Demonstrating feasibility in mental stress assessment using sensor fusion. IEEE Sensors Journal. 21(13), 14503–14514 (2020)
33. Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.A.: NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior research methods. 53(4), 1689–1696 (2021)

34. Hamilton, P..: Open source ECG analysis. In: Computers in cardiology, pp. 101–104. IEEE (2002)
35. Zhao, Z., Zhang, Y.: SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. Front. Physiol. 9, 727 (2018)
36. Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option?. Psicothema.
37. Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust?. Methodology.
38. McKinney, W.: Data structures for statistical computing in Python. In: van der Walt, S., Millman, J. (eds.) Proceedings of the 9th Python in Science Conference, pp. 51–56 (2010)
39. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Oliphant, T.E.: Array programming with NumPy. Nature. 585, 357–362 (2020)
40. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., SciPy 1.0 Contributors: SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature. Methods. 17(3), 261–272 (2020)
41. Damoun, N., Amekran, Y., Taiek, N., & El Hangouche, A. J. (2024). Heart rate variability measurement and influencing factors: Towards the standardization of methodology. Global cardiology science & practice, 2024(4), e202435.

The Autism-IT Linkage: A NeuroIS Research Agenda for Sociocognitive and Auditory Processing

Ronnie Jia[1], and Heather H. Jia[2]

[1] University of North Texas, Denton, Texas, USA
RonnieJia@gmail.com
[2] Illinois State University, Normal, Illinois, USA
hhjia@ilstu.edu

Abstract. Recent behavioral IS research has examined the autism-IT linkage, showing that autistic traits predict intrinsic interest in IT and are elevated among IT professionals and students. Evidence also suggests that many individuals in the IT field underperform in sociocognitive tasks and experience difficulty with auditory distractions. However, the underlying neurophysiological mechanisms remain poorly understood. This paper integrates emerging IS scholarship on autistic traits with neuroscience-based methods and proposes two directions for future neuroIS research. The first examines social cognition through theory of mind processing using eye tracking and pupillometry. The second investigates auditory processing using electroencephalography (EEG) to assess cognitive load and sensory gating. Together, these directions highlight how neuroIS can uncover the mechanisms underlying the autism-IT linkage and inform more effective workplace practices and IT pedagogy.

Keywords: Autistic Traits · Autism · Theory of Mind · Auditory Processing · EEG · Eye Tracking · Neurodiversity.

Introduction

Autism has long been characterized in the popular press as an "engineer's disorder" [29] and an "open secret" of the IT profession [22]. However, systematic examinations of the autism-IT relationship have only recently begun in IS research. Recent behavioral IS work suggests that autistic traits predict one's intrinsic interest in technology [16] and that those high in autistic traits are disproportionately represented in the IT profession [14].

Despite these advances, existing IS research has relied primarily on self-report and behavioral measures. As a result, the underlying neurophysiological mechanisms by which autistic traits shape sociocognitive and auditory processing in IT professionals remain poorly understood. Autism is a neurodevelopmental condition with well-documented neural correlates in social cognition and sensory processing—domains central to both IT work and IS inquiry. By integrating neuroscience methods with IS theory, neuroIS research is uniquely positioned to address this gap, moving beyond behavioral evidence to explain the underlying mechanisms of the phenomenon.

This paper synthesizes existing IS work on autistic traits and poses two high-level questions: 1) how do autistic traits in IT professionals shape their sociocognitive and sensory processing, and 2) what implications do these neurocognitive differences have for communication, collaboration, and alignment in IT contexts?

Guided by this agenda, we outline two directions for future neuroIS research: a) using eye tracking and pupillometry to examine social cognition, particularly theory of mind processing, and b) using electroencephalography (EEG) to investigate auditory processing differences. Together, these directions illustrate how neuroIS can advance theoretical understanding, methodological approaches, and practical insight into the autism–IT linkage.

Autism and Autistic Traits in IS Research

Autism is a neurodevelopmental condition characterized by persistent deficits in social communication and interaction, along with restricted or repetitive patterns of behavior, interests, or activities [2]. Despite its binary clinical diagnosis (positive/negative), autism is a spectrum condition, and autistic traits are continuously distributed throughout the whole population, with most people at subclinical levels [9]—one can be "a bit autistic" [13, p. 223]. Consequently, autistic traits can be examined as individual differences, allowing researchers to study meaningful variation regardless of diagnostic status.

Behavioral IS research has adopted this continuous view using self-report instruments such as the *Autism Spectrum Quotient* (AQ-50; [8]) and its brief versions (e.g., AQ-9; [15]). This line of work shows that autistic traits predict one's intrinsic interest in IT [16]. Further, IT workers exhibiting higher rates of clinically significant autistic traits and greater prevalence of autism diagnoses in their immediate families compared to the general population [14]. These results suggest that autistic traits may play a role in the attraction and self-selection into the IT field. As autism is associated with a range of differences in social cognition and sensory processing, these behavioral findings have important implications for IT professionals' sociocognitive and neurocognitive processing, particularly in light of widely observed social challenges among IT professionals [14]. NeuroIS methods offer a promising avenue for addressing these questions and informing this area of IS research.

Direction 1: Social Cognition, Theory of Mind, and Eye Tracking

As a core component of social cognition [1], *theory of mind* (ToM), or mentalizing, refers to the ability to attribute mental states—such as beliefs, intentions, and emotions—to oneself and others [26]. It enables one to interpret and predict others' behavior and is foundational to effective communication and collaboration. ToM impairments are a leading cognitive explanation for social communication differences in autism [4] [6].

Functional neuroimaging studies show that autistic people exhibit reduced activation in key ToM-related regions, including the medial prefrontal cortex (mPFC), the posterior superior temporal sulcus (pSTS), the precuneus, and the amygdala, leading to greater cognitive effort and difficulty [28]. Importantly, such differences are also observed in those high in autistic traits who do not meet diagnostic criteria [5] [7].

Cognitive Load

Recent IS research [14] provides converging behavioral evidence, showing that IT students required longer response times (RT) than management students in an advanced ToM task, called the *Reading the Mind in the Eyes Test* [7], which asks participants to infer mental states from photographs of the eye region of faces. Since RT is positively related to task complexity or cognitive load [30], longer RT suggests reliance on deliberate, effortful reasoning rather than intuitive mentalizing [20] [27]. However, neuroIS methods allow more direct assessment of these processes. For example, pupillometry provides a well-established physiological indicator of cognitive effort, with greater pupil dilation reflecting increased mental load during complex social inference tasks.

P1: Compared to the general population, IT professionals will exhibit greater cognitive effort in ToM tasks, reflected in greater pupil dilation.

Gaze Fixation Pattern

Beyond effort, eye tracking can reveal how visual attention is allocated to socially relevant cues. A substantial body of work shows that autism is associated with atypical gaze fixation patterns during social tasks, such as reduced fixation on eye regions, altered scanning of facial features, and different transitions between social and non-social elements of visual scenes [19]. Meta-analytic and systematic reviews confirm that reduced eye fixation is one of the most robust eye-tracking markers associated with autistic traits [24, 25]. Notably, even when individuals high in autistic traits perform adequately on explicit ToM tasks, eye-tracking data reveal reduced spontaneous attention to socially salient cues. Instead, they rely on effortful compensatory strategies that are mentally exhausting over time [20] [27].

P2: Compared to the general population, IT professionals will exhibit atypical gaze fixation patterns during ToM tasks, including reduced fixation on socially relevant cues and altered scan paths across social stimuli.

Methodological Considerations

The above propositions related to ToM processing can be examined using eye tracking and pupillometry. College students majoring in IT and business—serving as proxies for future IT professionals and their business counterparts—can be recruited as participants. Relevant demographic characteristics (e.g., gender, age, diagnostic status) should be controlled.

With respect to stimuli, P1 can be tested using the same social stimuli employed in prior behavioral studies, such as photographs of the eye region of faces from the *Reading the Mind in the Eyes Test* [7]. In contrast, examining fixation and scanning patterns for P2 would require different stimuli—those that present whole faces and include socially irrelevant cues—such as the *Cambridge Mindreading Face-Voice Battery* [12].

Direction 2: Auditory Processing and EEG

In addition to social cognition, autism is associated with altered sensory processing, including hyper- and hypo-responsiveness to sensory input [2]. Across sensory domains, auditory processing differences show the strongest association with autistic traits in both clinical and nonclinical populations [3]. These differences are particularly relevant for IT education and work contexts, where communication often relies heavily on spoken language in lectures, meetings, and collaborative work.

Auditory Cognitive Load

Although practitioner accounts suggest that many IT professionals describe themselves as "visual" thinkers [18] [21], such self-descriptions may reflect not only visual strengths but also relative weakness in auditory processing. Anecdotal classroom observations similarly indicate that many IT students disengage quickly from spoken lectures and shift toward visual or hands-on materials, even when interest in the content remains high. From a sensory processing perspective, such disengagement is consistent with elevated auditory cognitive load rather than lack of motivation or simple preference.

If auditory processing is more demanding for IT students and professionals, this has important implications for classroom instruction, workplace meetings, and media choice. Traditional theories such as media richness theory emphasize matching message equivocality to medium richness [10]. A sensory-processing account instead highlights the importance of person-technology fit, where communication effectiveness also depends on one's capacity to process and integrate sensory input. EEG provides a real-time measure of cognitive demands during auditory processing. Increased power in the alpha (8–12 Hz) and theta (4–8 Hz) frequency bands is indicative of heightened cognitive load, attentional control, and working memory engagement during challenging listening tasks.

P3: Compared to the general population, IT professionals will experience greater cognitive load during auditory processing tasks, reflected in increased alpha and theta EEG activity.

Auditory Sensory Gating

Auditory cognitive load is closely related to—but distinct from—*auditory sensory gating*, which refers to the neural system's ability to suppress responses to repetitive or irrelevant auditory stimuli before they consume higher-order cognitive resources. By inhibiting redundant sensory input at early stages of processing, effective gating protects attention and reduces the likelihood of sensory overload [23].

Research shows that reduced auditory gating leads to attentional difficulties and cognitive overload and is frequently observed in autistic individuals. Importantly, impaired gating can amplify downstream cognitive load by allowing irrelevant auditory input to compete for cognitive resources, thereby increasing attentional demands and task effort [23].

In IT education and work contexts, reduced gating could manifest as higher distractibility in sound-rich environments, difficulty maintaining focus during meetings, and preferences for quiet or acoustically controlled work settings. Accordingly, implementing collaborative workspaces by collocating team members in a single shared space—an approach that may be effective for workers in other occupations—may prove counterproductive for IT employees. These differences help explain well-documented preferences among many IT professionals for environments that minimize auditory demands. For example, individuals high in autistic traits report higher cognitive absorption in task performance but greater sensitivity to noise and distractions [17]. Unsurprisingly, IT workers have long been interested in telecommuting or "any alternative which would … reduce work interruptions" [11, p. 138]. During the Covid-19 pandemic, as offices became quieter, some IT workers who previously preferred working from home even returned to the office due to reduced noise and distraction [31].

P4: Compared to the general population, IT professionals will exhibit reduced auditory sensory gating, reflected in diminished suppression of repetitive stimuli (S2 relative to S1) and reduced early ERP responses such as the P1 component.

Methodological Considerations

Similar to the earlier propositions, P3 and P4 can be examined using college students majoring in IT and non-IT fields as proxies for IT professionals and the general population, respectively, while controlling for relevant demographic characteristics.

With respect to stimuli, P3 can be tested using both speech and non-speech auditory tasks. Speech stimuli may include spoken words or sentences presented under varying levels of complexity or speech-to-noise ratio, whereas non-speech stimuli (e.g., pure tones) allow assessment of more basic auditory processing demands independent of linguistic processing. Cognitive load during these tasks is reflected in alpha and theta EEG activity.

P4 can be assessed using the standard paired-click paradigm, where pairs of identical clicks are presented approximately 500ms apart, and gating is indexed by the degree of neural response suppression to the second stimulus (S2) relative to the first (S1). This suppression effect is typically quantified using early ERP components such as P1, N1, and P2, with reduced S2/S1 suppression indicating impaired filtering of redundant auditory input [23].

Conclusion

This paper advances neuroIS research by integrating IS scholarship on autistic traits with neuroscience methods to examine the neurocognitive mechanisms underlying sociocognitive and auditory processing in IT populations. Using eye tracking, pupillometry, and EEG, the proposed directions move the autism-IT linkage beyond self-report and behavioral outcomes toward direct assessment of underlying mechanisms.

Key contributions lie in extending core IS theories, such as media richness and alignment, through a sensory-processing and neurodiversity lens. Differences in theory-of-mind and auditory processing suggest that business-IT communication challenges may stem from systematic variation in sociocognitive and sensory processing patterns rather than solely from incentives, attitudes, or knowledge gaps. This view highlights person-job-technology fit as an important boundary condition for coordination and collaboration in IT contexts.

Building on this theoretical reorientation, future design research can investigate how communication technologies, collaboration practices, work environments, and pedagogical approaches can be structured to manage cognitive and auditory load while maintaining effectiveness. While this paper outlines only two illustrative directions, it demonstrates the broader potential of neuroIS to inform both IS theory and evidence-based improvements in IT work practices and IT pedagogy.

Acknowledgements. RJ gratefully acknowledges partial support from an appointment with the U.S. Department of Homeland Security (DHS) Science & Technology (S&T) Directorate Office of University Programs Summer Research Team Program for Minority Serving Institutions, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by ORAU under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of DHS, DOE, or ORAU/ORISE.

References

1. Adolphs, R.: Recognizing emotion from facial expressions: psychological and neurological mechanisms. Behavioral and Cognitive Neuroscience Reviews, vol. 1(1), pp. 21-62. (2002)
2. American Psychiatric Association (APA): Diagnostic and Statistical Manual of Mental Disorders (5th ed.). DSM-5. Washington, DC (2013)
3. Bang, P., & Igelström, K. Modality-specific associations between sensory differences and autistic traits. Autism, vol. 27(7), pp. 2158–2172. (2023)
4. Baron-Cohen, S.: Autism: A specific cognitive disorder of "mind-blindness', International Review of Psychiatry, vol. 2(1), pp. 81-90, (1990)
5. Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M.: Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger Syndrome. Journal of Child Psychology & Psychiatry & Allied Disciplines, vol. 38(7), pp. 813–822, (1997)
6. Baron-Cohen, S., Leslie, A. M., & Frith, U.: Does the autistic child have a "theory of mind"? Cognition, vol. 21(1), pp. 37–46, (1985)
7. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I.: The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. Journal of Child Psychology & Psychiatry, vol. 42(2), pp. 241-251, (2001)
8. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E.: The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism & Developmental Disorders, vol. 31(1), pp. 5-17, (2001)
9. Constantino, J.N., & Todd, R.D.: Autistic traits in the general population: A twin study. Archives of General Psychiatry, vol. 60(5), pp. 524–530, (2003)
10. Daft, R.L. & Lengel, R.H. Organizational information requirements, media richness and structural design. Management Science, 32(5), pp. 554-571, (1986)
11. DeSanctis, G.: Attitudes toward telecommuting: Implications for work-at-home programs, Information & Management, vol. 7(3), pp. 133-139, (1984)
12. Golan, O., Baron-Cohen, S., & Hill, J.: The Cambridge mindreading (CAM) face-voice battery: Testing complex emotion recognition in adults with and without Asperger syndrome. Journal of Autism and Developmental Disorders, vol. 36(2), 169–183, (2006)
13. Happé, F., & Frith, U.: Annual research review: Looking back to look forward–Changes in the concept of autism and implications for future research. Journal of Child Psychology & Psychiatry, vol. 61(3), pp. 218-232, (2020)

14. Jia, R., Riemenschneider, C. & Jia, H.H.: Autistic traits, theory of mind, and implications for IT professionals' emotional labor and mental wellbeing, MIS Quarterly, (2026) https://doi.org/10.25300/MISQ/2025/19395
15. Jia, R., Steelman, Z.R., & Jia, H.H.: Psychometric assessments of three self-report autism scales (AQ, RBQ-2A, and SQ) for general adult populations. Journal of Autism & Developmental Disorders, vol. 49(5), pp. 1949-1965, (2019)
16. Jia, R., Steelman, Z.R., & Jia, H.H.: What makes one intrinsically interested in IT? An exploratory study on influences of autistic tendency and gender in the U.S. and India. MIS Quarterly, vol. 46(3), 1603-1634, (2022)
17. Jia, R., Steelman, Z.R. & Jia, H.H. An Integration of Cognitive Absorption and Interruption Overload Through the Lens of Neurodiversity: Some Preliminary Results. Proceedings of Hawaii International Conference on System Sciences, (2023)
18. Kalidasan, K.: Think in shapes: A visual thinking guide for engineers. LinkedIn, April 14, (2025). https://www.linkedin.com/pulse/think-shapes-visual-thinking-guide-engineers-kannan-kalidasan-zxnje
19. Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D.: Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry,* vol. *59*(9), pp. 809–816, (2002)
20. Livingston, L. A., Colvert, E.: Social Relationships Study Team, Bolton, P., & Happé, F.: Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. Journal of Child Psychology & Psychiatry, vol. 60(1), pp. 102–110, (2019)
21. Mallinson, A.: Visual thinking for engineers. MIT News, September 10, (2014). https://news.mit.edu/2014/visual-thinking-engineers
22. Mayor, T. IT's open secret. ComputerWorld, May 5. pp. 23-27. (2008)
23. Miller, S. E., Graham, J., & Schafer, E.: Auditory sensory gating of speech and nonspeech stimuli. Journal of Speech, Language, and Hearing Research, vol. 64(4), pp. 1404–1412, (2021)
24. Nayar, K., Voyles, A.C., Kiorpes, L., & Di Martino, A.: Atypical visual attention to faces in autism spectrum disorder: A meta-analysis of eye-tracking studies. *Molecular Autism,* vol. *13*, Article 21, (2022)
25. Papagiannopoulou, E.A., Chitty, K.M., Hermens, D.F., Hickie, I.B., & Lagopoulos, J.: A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience,* vol. *9*(6), pp. 610–632, (2014)
26. Premack, D., & Woodruff, G.: Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences, vol. 1(4), pp. 515-526, (1978)
27. Senju, A., et al.: Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. Science, vol. 325(5942), pp. 883–885, (2009)
28. Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Sallet, J., & Kanske, P.: Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychological Bulletin, vol. 147(3), pp. 293–327, (2021)
29. Silberman, S. (2001). The geek syndrome. Wired. https://www.wired.com/2001/12/aspergers/
30. Turcotte, J., & Oddson, B.: Managing Increased Cognitive Load in a Guided Search. Perceptual and Motor Skills, vol. 129(3), pp. 378-398, (2022)
31. Uddin, G.; Alam, O. & Serebrenik, A.: A qualitative study of developers' discussions of their problems and joys during the early COVID-19 months. Empirical Software Engineering, vol. 27, (2022)

Reflective and Impulsive GenAI Use in Workplace Tasks: Proposal for an Eye-Tracking Study of Adults with ADHD

Theresa Wortmann[1] and Julia Sarah Kroenung[1]

[1] Chair of Sociotechnical Information Systems Design, FernUniversität in Hagen, Hagen, Germany
{theresa.wortmann,julia.kroenung}@fernuni-hagen.de

Abstract. Generative artificial intelligence (GenAI), particularly large language model (LLM)-based assistants, is increasingly embedded in knowledge work, requiring users to regulate attention and coordinate multiple information sources during task execution. At the same time, attention-deficit/hyperactivity disorder (ADHD), affecting approximately 2-5% of adults worldwide, is associated with systematic variability in executive functioning. Yet it remains unclear how variability in executive function shapes attention allocation and interaction dynamics during human-GenAI collaboration. Drawing on the Reflective-Impulsive Model, this study proposes a controlled eye-tracking experiment comparing adults with ADHD and non-ADHD controls in LLM-supported knowledge work. Using eye-tracking metrics, we examine process-level differences in attentional regulation between reflective and impulsive LMM-based GenAI use. The study contributes to Neuro-Information Systems research by operationalizing reflective and impulsive GenAI use through gaze-based indicators and examining ADHD-related variability in executive functioning as a boundary condition in LLM-supported knowledge work.

Keywords: Eye-Tracking · GenAI · Knowledge Work · ADHD · Reflective-Impulsive Model

Introduction

Generative Artificial Intelligence (GenAI) is rapidly becoming part of everyday knowledge work [1]. Particularly, large language model (LLM)-based assistants are increasingly relevant because they support language-intensive tasks such as drafting texts, structuring information, problem-solving, and ideation that are central to many contemporary professional activities [1, 2]. Global survey data indicate that approximately 75% of employees report already using AI at work, with nearly half having started using Artificial Intelligence (AI) within the last six months [3], suggesting that LLM-supported task execution is no longer emerging but increasingly normative in digital work practices. Employees now routinely interact with AI systems while performing cognitively demanding tasks across a wide range of digital work contexts [2, 4]. At the same time, attention-deficit/hyperactivity disorder (ADHD) affects approximately 2-5% of adults worldwide [5, 6], representing a substantial population engaged in digital work environments.

Knowledge work is inherently attention- and control-intensive, requiring sustained goal maintenance, inhibition of distractions, and integration of information across multiple sources [7, 8]. LLM-supported tasks further amplify these demands, as users must continuously coordinate task requirements, source material, AI-generated output, and their own responses [9]. ADHD is associated with systematic variability in executive functions, particularly sustained attention, inhibitory control, and working memory [10]. These mechanisms are central to regulating behavior in such cognitively demanding, multi-channel task environments in digital knowledge work [11]. Despite this theoretical relevance, little is known about how adults with ADHD engage with LLM-based GenAI during complex workplace tasks.

To conceptualize variability in LLM-supported behavior, we draw on the Reflective-Impulsive Model (RIM), which posits that behavior arises from the interplay between a deliberative, reflective system and a faster, cue-driven, impulsive system [12]. Yet, LLM-supported work has rarely been examined at a behavioral dual-process level, particularly in populations characterized by variability in executive functioning, such as adults with ADHD. By capturing visual attention allocation and processing trajectories, eye-tracking enables direct observation of attentional regulation, information integration, and processing stability during task execution [13]. Prior research demonstrates that eye-movement metrics are sensitive to executive-function differences and can reveal systematic process-level variability even when final performance outcomes are comparable [14]. Against this background, our study proposal addresses the following research questions:

RQ1: How do visual attention and eye-movement patterns during LLM-supported workplace tasks differ between adults with ADHD and non-ADHD controls?

RQ2: How does LLM-based GenAI use differ in reflective versus impulsive processing between adults with ADHD and non-ADHD controls?

Theoretical Background and Related Work

Reflective Impulsive Model. The Reflective Impulsive Model (RIM) conceptualizes behavior as the joint outcome of two interactive processing systems: a reflective system that relies on deliberative, rule-based reasoning like facts, values, and explicit goals, and an impulsive system that operates through associative links and motivational orientations, eliciting faster, cue-driven responses [12]. The reflective system is resource-demanding and relies on executive control, whereas the impulsive system operates automatically and requires minimal cognitive capacity. Under conditions of distraction, arousal, or high cognitive load, reflective processing can become unstable, increasing the influence of impulsive responses [12].

In the context of information technologies (IT), this perspective has been extended into a triadic model, in which technology use is driven by reflective and impulsive systems, modulated by an interoceptive awareness system that generates temptations [15]. Interoceptive signals, such as cravings and boredom, affect decision-making by weakening reflective control and strengthening impulsive influences [15]. Further studies link impulsivity, executive dysfunction, and specific inhibitory control deficits to disordered online behavior, with attentional impulsivity playing a central role and interacting with weaker executive functions and inhibitory control [16]. Hence, reflective processes are largely dependent on the prefrontal cortex, impulsive processes on the amygdala-striatum, and interoceptive awareness on the insula [15]. This mapping is relevant for ADHD because ADHD has been associated with differences in neurocognitive systems that support executive control and impulse regulation, which are central to maintaining reflective processing under cognitive demands [17].

ADHD, GenAI, and Knowledge Work. Knowledge work requires sustained goal maintenance, selective attention, inhibition of irrelevant information, and integration of multiple information sources [18]. Many professional tasks, such as writing reports, structuring information, evaluating alternatives, and generating ideas, depend on executive functions that regulate attention, working memory, and strategic control [7, 19]. These tasks are increasingly shaped by LLM-based GenAI assistants, which are embedded in interactive, prompt-based workflows and support a broad range of language-intensive cognitive tasks. As a result, these systems are reshaping how knowledge is generated, organized, evaluated, and applied in contemporary work settings [1, 20-23]. At the same time, effective use of LLM-based GenAI places additional self-regulatory demands on users. Rather than processing a single information source, users must continuously coordinate attention across task instructions, source materials, AI-generated output, and their own evolving responses. Consequently, successful use depends not only on what the system produces, but also on how individuals monitor, verify, and strategically integrate LLM-generated content during task execution [1, 24, 25].

These demands are relevant for adults with ADHD, which is a neurodevelopmental condition characterized by systematic variability in executive functioning [10, 26]. ADHD-related performance differences tend to become more pronounced in demanding tasks that require individuals to continuously ignore irrelevant information and repeatedly shift their attention between task elements. When inhibitory resources are strained, regulating responses becomes more difficult, increasing susceptibility to more impulsive, less reflective responding [27]. At the same time, individuals with ADHD may experience episodes of sustained, high-intensity engagement under motivating conditions (“hyperfocus”), which can support strong goal-directed processing in specific task contexts and may reflect strong reflective processing [28, 29].

The effect of LLM-supported knowledge work on adults with ADHD can therefore be double-edged: While LLM-based GenAI output can externalize structure and reduce working-memory load, it can also introduce salient cues that compete for limited executive resources [9]. Although digital technologies have been discussed as potential training or support tools for executive functioning [30], empirical evidence regarding GenAI-assisted work in adult ADHD is limited and heterogeneous [7]. Understanding this variability requires a process-oriented perspective on situated human-AI interaction [31].

Research Propositions

To the best of our knowledge, no prior Neuro Information Systems (NeuroIS) study has systematically examined how adults with ADHD engage with LLM-supported workplace tasks using eye-tracking measures. Existing literature provides observations regarding the cognitive demands of LLM-supported work [9, 21], executive-function variability associated with ADHD [6, 26], and the usefulness of gaze metrics for capturing latent attentional processes [32, 33]. These observations can be linked to the RIM and guide directional expectations for the proposed study.

LLM-supported workplace tasks require users to coordinate task instructions, source material, generated output, and the production of responses [21, 25, 31]. Such environments place demands on sustained attention, inhibitory control, and structured information integration [11, 18]. Prior research on ADHD suggests that cognitively demanding multi-source tasks may be associated with different attentional regulation patterns and greater variability in executive control [6, 26, 34]. Based on these observations, we derive the following propositions in the context of different workplace tasks:

Proposition 1: During structuring tasks, adults with ADHD are expected to display stronger impulsive processing than non-ADHD controls, reflected in less stable gaze patterns, faster shifts between interface elements, and reduced sustained attention to task instructions and source material.

Proposition 2: During evaluation tasks, adults with ADHD are expected to differ from non-ADHD controls in reflective processing, shown in fewer verification-oriented gaze transitions between task requirements, LLM output, and response areas under high executive-control demands.

Proposition 3: During creative tasks, adults with ADHD are expected to exhibit sustained and structured attention patterns stronger than non-ADHD controls, indicating reflective and goal-directed engagement under stimulating task conditions.

Proposition 4: Across workplace task types, adults with ADHD are expected to show stronger shifts between reflective and impulsive processing than non-ADHD controls, such that structuring and evaluation tasks more strongly elicit impulsive tendencies, whereas creative tasks more strongly elicit reflective and sustained engagement.

Study Proposal

Proposed Study Design. The study aims to examine how adults with ADHD differ from non-ADHD controls in their use of LLM-based GenAI during workplace-relevant knowledge tasks. Building on the theoretical background, we first conduct a qualitative pre-study to validate a set of realistic task scenarios that preserve ecological validity while enabling experimental control. This step follows NeuroIS and eye-tracking research emphasizing the importance of theory-driven and standardized stimuli for process measurement [33, 35].

Based on the literature on knowledge work and cognitive task demands, the pre-study focuses on three workplace task types that differ systematically in their processing requirements while avoiding specialized domain knowledge: structuring tasks, which require organizing and synthesizing information into a coherent format [36], creative tasks, which require idea generation and divergent thinking [21, 23] and evaluation tasks, which require reviewing, comparing, or critically assessing AI-generated content [21, 36]. These three task types were selected because they vary in levels of ambiguity, verification demands, and executive control requirements, making them suitable contexts for examining reflective and impulsive LLM-based GenAI use [37, 38]. Each task is embedded in a standardized digital interface containing four predefined Areas of Interest (AOIs): task instructions, source material, LLM output, and response area. This common interface structure supports comparability across conditions and enables consistent process-level gaze analysis [39]. A pilot validation sample will assess scenario realism, clarity, perceived difficulty, and the appropriateness of LLM support. Based on this feedback, task wording, timing, and AOI boundaries will be refined before the main study.

The empirical investigation follows a two-step design. First, an online experiment with university students will examine behavioral differences in LLM-supported task performance across the three task types using a larger sample. Participants from the ADHD group will be classified based on a formal ADHD diagnosis and a validated adult ADHD self-report measurement [40, 41]. Remaining participants will form the non-AHD control group. This first step allows adequately powered inferential analyses and strengthens group classification validity. Second, a controlled lab-based eye-tracking experiment with a smaller subsample will investigate the underlying attentional and cognitive mechanisms under standardized conditions. We propose a mixed experimental design with a between-subjects factor (ADHD vs. non-ADHD control group) and a within-subjects factor (three workplace task types: structuring, creative, and evaluation). To strengthen internal validity, all participants complete the same three tasks using the same standardized LLM-based assistant interface, while the order of tasks is randomized across participants to reduce sequence, fatigue, and learning effects [38, 39]. Participants will be instructed to meaningfully engage with the LLM assistant during each task, while retaining autonomy regarding the extent of use. The participant compensation will be independent of task performance. Task instructions, exposure time, and presented LLM outputs are standardized wherever possible. This design enables us to examine whether adults with ADHD differ from controls in task behavior, visual attention allocation, gaze transitions, and indicators of reflective versus impulsive LLM-based GenAI use across distinct workplace task contexts. At the same time, the comparison across task types allows an analysis of whether certain task demands increase reliance on impulsive, cue-driven processing or foster more stable, reflective, and goal-directed engagement.

Eye-Tracking Measurements. Eye-tracking is a non-intrusive process-tracking method that captures visual attention and information processing during task execution by recording where, when, and for how long individuals allocate gaze to elements of a digital interface [33, 42, 43]. To capture reflective and impulsive LLM-based GenAI use as process behavior in workplace tasks, we follow established eye-tracking practices in IS research, which emphasize theory-driven AOIs, fixation- and transition-based metrics, and triangulation with behavioral data [e.g. 14, 33, 44-47].

Prior eye-tracking studies suggest that adults with ADHD often achieve comparable task accuracy while exhibiting higher cognitive processing effort and less stable attentional allocation [48]. Moreover, adults with ADHD tend to show shorter fixation durations on task-relevant information, increased saccadic activity, and less efficient or more irregular gaze trajectories, particularly in tasks requiring executive control [34, 49]. These patterns reflect differences in attentional stability and visual information integration during cognitively demanding tasks. Accordingly, we operationalize attentional regulation using three gaze indicators: First, fixation duration and total dwell time on task-relevant AOIs capture sustained visual engagement with relevant task elements. Second, saccadic activity reflects the frequency of gaze shifts and thus indicates the degree of visual exploration and attentional reorientation during task processing. And third, gaze dispersion captures the spatial variability of gaze distribution across the interface and serves as an indicator of fragmented or unstable visual exploration under cognitive load [42, 50, 51].

Building on the RIM, we examine gaze patterns consistent with reflective versus impulsive LLM-based GenAI use [12, 15]. Reflective use is expected to be associated with longer dwell times on task requirements and constraints, more frequent regressions and transitions between instructions, LLM output, and the participant's draft, and extended engagement with LLM output prior to finalizing the response [32, 52]. In contrast, impulsive use can be reflected in shorter reading phases of LLM output, rapid gaze shifts toward response production with limited revisiting of task constraints, and more dispersed gaze patterns during demanding trials [32, 53]. Because eye-tracking captures visual attention rather than cognitive processes directly, gaze metrics are interpreted as indicators of underlying reflective or impulsive processing.

Hence, ADHD-related differences in executive control can affect the stability of reflective processing so that these gaze-based indicators allow us to examine whether GenAI-supported task behavior differs systematically between ADHD and non-ADHD participants. To strengthen inference, gaze metrics are triangulated with task performance and GenAI interaction logs, following NeuroIS recommendations for multi-method measurement [33, 45]. Integrating gaze and behavioral features further supports the identification of meaningful individual differences in ADHD-related processing patterns [49, 53].

Expected Results and Contributions

Based on the theoretical framework, we expect systematic group differences in process-level attention dynamics during LLM-supported workplace tasks. In line with prior research showing that comparable task outcomes can mask differences in cognitive effort and attentional stability [48, 49], we anticipate that adults with ADHD will show greater variability in attentional stability, particularly in cognitively demanding or ambiguous tasks. At the same time, we expect that specific task contexts may elicit sustained, structured engagement patterns consistent with hyperfocus, reflecting strong goal-directed processing under motivating conditions. Rather than assuming uniform deficits, the study therefore examines variability in how reflective and impulsive dynamics unfold during human-GenAI collaboration. Reflective use is expected to appear in stable engagement with task requirements and structured integration of LLM output, whereas impulsive use may involve rapid acceptance patterns and less iterative verification. We further expect task characteristics to moderate these dynamics by increasing the demands of executive functioning.

Theoretically, this research contributes to NeuroIS research in three ways. First, it advances a process-level understanding of LLM-supported knowledge work by linking the RIM to observable human-GenAI interaction, demonstrating how reflective and impulsive mechanisms can be operationalized through gaze-based process indicators. Second, by conceptualizing ADHD not as a deficit category but as a form of systematic executive-function variability, the study positions ADHD as a meaningful boundary condition for understanding heterogeneous AI use behavior, highlighting both regulatory challenges and context-dependent strengths. Third, it demonstrates how theory-driven eye-tracking metrics capture latent processing dynamics even when final performance appears comparable.

Practically, this research contributes to the design of inclusive LLM-supported work systems. By identifying how different attentional and interaction patterns unfold across cognitive profiles, the findings can inform interface structures, verification cues, and workflow designs that support stable and reflective LLM-based GenAI engagement. Importantly, the study moves beyond a deficit-oriented framing of ADHD. Instead of positioning ADHD as a problem to be corrected, the findings may help organizations better understand when LLMs can function as a cognitive support and when task configurations may increase attentional strain. Such insights can guide the development of adaptive human-GenAI collaboration strategies that leverage cognitive strengths while mitigating context-specific challenges. In doing so, the research supports more nuanced and inclusive integration of LLM-based GenAI into organizational knowledge work.

Finally, this study proposal has several limitations that should be acknowledged and offers avenues for future research. First, the planned recruitment of university students as proxies for workplace knowledge workers may limit ecological validity. Although student samples are common in experimental research and suitable for identifying underlying cognitive mechanisms, students may differ from employed adults in work experience, organizational routines, and habitual LLM-based GenAI use. This may also apply to adults with ADHD, whose self-regulation demands can vary across academic and workplace contexts. Future research should therefore extend the design to working populations and organizational settings. Second, this study focuses on LLM-based GenAI assistants as a specific and currently dominant form of GenAI in knowledge work. While this focus allows for a controlled and theoretically grounded investigation of human-GenAI interaction, it also limits the generalizability of the findings to other types of generative systems like image-based, multimodal, or agentic AI systems. Different GenAI technologies may impose distinct cognitive demands and interaction patterns, which are not captured in the present design. Future research should therefore extend this work by examining and comparing multiple forms of GenAI to better understand how different system characteristics shape attention, cognitive processing, and user behavior in knowledge work contexts. Third, although controlled task scenarios improve internal validity, they necessarily simplify real workplace environments. In addition, characteristics such as the length or complexity of LLM-generated responses may influence gaze behavior. Future research may therefore complement controlled experiments with more naturalistic settings and additional controls for output characteristics.

References

1. Benbya, H., Strich, F., Tamm, T.: Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. Journal of the Association for Information Systems 25, 23-36 (2024)
2. Bankins, S., Ocampo, A.C., Marrone, M., Restubog, S.L.D., Woo, S.E.: A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. Journal of Organizational Behavior 45, 159-182 (2023)
3. Statista: User rate of generative artificial intelligence (AI) in the workplace globally in 2024. Website Name. https://www.statista.com/statistics/1482102/rate-of-generative-ai-utilization-globally/ (Accessed 2026-02-26)
4. Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R.: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. SSRN Electronic Journal, (2023)
5. Simon, V., Czobor, P., Balint, S., Meszaros, A., Bitter, I.: Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. Br J Psychiatry 194, 204-11 (2009)
6. Ayano, G., Tsegay, L., Gizachew, Y., Necho, M., Yohannes, K., Abraha, M., Demelash, S., Anbesaw, T., Alati, R.: Prevalence of attention deficit hyperactivity disorder in adults: Umbrella review of evidence generated across the globe. Psychiatry Res 328, 115449 (2023)
7. Pergantis, P., Bamicha, V., Skianis, C., Drigas, A.: AI Chatbots and Cognitive Control: Enhancing Executive Functions Through Chatbot Interactions: A Systematic Review. Brain Sci 15, (2025)
8. Diamond, A.: The Early Development of Executive Functions. In: Bialystok, E., Craik, F. (eds.) Lifespan Cognition: Mechanisms of change. Vol. Oxford University Press, (2006)
9. Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A.E., Sarkar, A., Sellen, A., Rintel, S.: The Metacognitive Demands and Opportunities of Generative AI. Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1-24. (2024)
10. Barkley, R.A.: Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 121, 65-94 (1997)
11. Rann, J.C., Almor, A.: An examination of sustained attention during complex multitasking scenarios. Cogn Res Princ Implic 10, 67 (2025)
12. Strack, F., Deutsch, R.: Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 8, 220-47 (2004)
13. Deng, R., Gao, Y.: A review of eye tracking research on video-based learning. Educ Inf Technol (Dordr) 28, 7671-7702 (2023)
14. Bera, P., Soffer, P., Parsons, J.: Using Eye Tracking to Expose Cognitive Processes in Understanding Conceptual Models1. MIS Quarterly 43, 1105-1126 (2019)
15. Turel, O., Bechara, A.: A Triadic Reflective-Impulsive-Interoceptive Awareness Model of General and Impulsive Information System Use: Behavioral Tests of Neuro-Cognitive Theory. Front Psychol 7, 601 (2016)
16. Wegmann, E., Muller, S.M., Turel, O., Brand, M.: Interactions of impulsivity, general executive functions, and specific inhibitory control explain symptoms of social-networks-use disorder: An experimental study. Sci Rep 10, 3866 (2020)

17. Cortese, S., Kelly, C., Chabernaud, C., Proal, E., Di Martino, A., Milham, M.P., Castellanos, F.X.: Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. Am J Psychiatry 169, 1038-55 (2012)
18. Kalgotra, P., Sharda, R., McHaney, R.: Don't Disturb Me! Understanding the Impact of Interruptions on Knowledge Work: an Exploratory Neuroimaging Study. Information Systems Frontiers 21, 1019-1030 (2017)
19. Sullivan, Y., Davis, F., Koh, C.: Executive Functions and Information Systems Learning. MIS Quarterly 46, 813-880 (2022)
20. Banh, L., Strobel, G.: Generative artificial intelligence. Electronic Markets 33, (2023)
21. Brachman, M., El-Ashry, A., Dugan, C., Geyer, W.: Current and Future Use of Large Language Models for Knowledge Work. Proceedings of the ACM on Human-Computer Interaction 9, 1-24 (2025)
22. Alavi, M., Leidner, D.E., Mousavi, R.: Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems 25, 1-12 (2024)
23. Feuerriegel, S., Hartmann, J., Janiesch, C., Zschech, P.: Generative AI. Business & Information Systems Engineering 66, 111-126 (2023)
24. Yan, L., Greiff, S., Teuber, Z., Gasevic, D.: Promises and challenges of generative artificial intelligence for human learning. Nat Hum Behav 8, 1839-1850 (2024)
25. Wu, S., Liu, Y., Ruan, M., Chen, S., Xie, X.Y.: Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Sci Rep 15, 15105 (2025)
26. Diamond, A.: Executive functions. Annu Rev Psychol 64, 135-68 (2013)
27. Senkowski, D., Ziegler, T., Singh, M., Heinz, A., He, J., Silk, T., Lorenz, R.C.: Assessing Inhibitory Control Deficits in Adult ADHD: A Systematic Review and Meta-analysis of the Stop-signal Task. Neuropsychol Rev 34, 548-567 (2024)
28. Ashinoff, B.K., Abu-Akel, A.: Hyperfocus: the forgotten frontier of attention. Psychol Res 85, 1-19 (2021)
29. Garcia Pimenta, M., Gruhnert, R.K., Fuermaier, A.B.M., Groen, Y.: The role of executive functions in mediating the relationship between adult ADHD symptoms and hyperfocus in university students. Res Dev Disabil 144, 104639 (2024)
30. Small, G.W., Lee, J., Kaufman, A., Jalil, J., Siddarth, P., Gaddipati, H., Moody, T.D., Bookheimer, S.Y.: Brain health consequences of digital technology use . Dialogues Clin Neurosci 22, 179-187 (2020)
31. Hao, X., Demir, E., Eyers, D.: Exploring collaborative decision-making: A quasi-experimental study of human and Generative AI interaction. Technology in Society 78, (2024)
32. Eckstein, M.K., Guerra-Carrillo, B., Miller Singley, A.T., Bunge, S.A.: Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? Dev Cogn Neurosci 25, 69-91 (2017)
33. Vasseur, A., Passalacqua, M., Sénécal, S., Léger, P.-M.: The Use of Eye-tracking in Information Systems Research: A Literature Review of the Last Decade. AIS Transactions on Human-Computer Interaction 15, 292-321 (2023)
34. Mohammadhasani, N., Capri, T., Nucita, A., Iannizzotto, G., Fabio, R.A.: Atypical Visual Scan Path Affects Remembering in ADHD. J Int Neuropsychol Soc 26, 557-566 (2020)
35. Balapour, A., Riedl, R.: Ecological Validity in NeuroIS Research: Theory, Evidence, and a Roadmap for Future Studies. Journal of the Association for Information Systems 26(1), 9-65 (2024)
36. Wang, A., Zhao, Y., Ma, F.: A diary study of information-intensive work tasks in the modern workplace: Investigating task descriptions and task processes. Journal of the Association for Information Science and Technology 77, 663-681 (2025)
37. Byström, K., Järvelin, K.: Task complexity affects information seeking and use. Information Processing & Management 31, 191-213 (1995)

38. Braver, T.S., Kizhner, A., Tang, R., Freund, M.C., Etzel, J.A.: The Dual Mechanisms of Cognitive Control Project. J Cogn Neurosci, 1-26 (2021)
39. Orquin, J.L., Ashby, N.J.S., Clarke, A.D.F.: Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. Journal of Behavioral Decision Making 29, 103-115 (2015)
40. Ustun, B., Adler, L.A., Rudin, C., Faraone, S.V., Spencer, T.J., Berglund, P., Gruber, M.J., Kessler, R.C.: The World Health Organization Adult Attention-Deficit/Hyperactivity Disorder Self-Report Screening Scale for DSM-5. JAMA Psychiatry 74, 520-527 (2017)
41. Kessler, R.C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M.J., Jin, R., Secnik, K., Spencer, T., Ustun, T.B., Walters, E.E.: The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med 35, 245-56 (2005)
42. Joseph, A.W., Murugesh, R.: Potential Eye Tracking Metrics and Indicators to Measure Cognitive Load in Human-Computer Interaction Research. Journal of scientific research 64, 168-175 (2020)
43. Just, M.A., Carpenter, P.A.: A theory of reading: From eye fixations to comprehension. Psychological Review 87, 329-354 (1980)
44. Pfeiffer, J., Pfeiffer, T., Meißner, M., Weiß, E.: 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 (2020)
45. 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., Brocke, J.v., Weber, B.: On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for NeuroIS. MIS Quarterly 36, 679-702 (2012)
46. Riedl, R., Fischer, T., Léger, P.-M., Davis, F.D.: A Decade of NeuroIS Research. ACM SIGMIS Database: the DATABASE for Advances in Information Systems 51, 13-54 (2020)
47. Ehmke, C., Wilson, S.: Identifying web usability problems from eye-tracking data. Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK, (2007)
48. Kyriacou, M., Rummelhoff, C., Koder, F.: Irony Processing in Adults with ADHD: Evidence From Eye-Tracking and Executive Attention Tasks. J Atten Disord 29, 724-744 (2025)
49. Chen, X., Wang, S., Yang, X., Yu, C., Ni, F., Yang, J., Tian, Y., Ye, J., Liu, H., Luo, R.: Utilizing artificial intelligence-based eye tracking technology for screening ADHD symptoms in children. Front Psychiatry 14, 1260031 (2023)
50. Shimojo, S., Simion, C., Shimojo, E., Scheier, C.: Gaze bias both reflects and influences preference. Nat Neurosci 6, 1317-22 (2003)
51. Alrefaei, D., Sankar, G., Djamasbi, S., Norouzi Nia, J., Strong, D.: Detecting Anxiety via Eye movements: a User Experience Approach to Research and Development. SIGHCI 2024 Proceedings 25, (2025)
52. Cao, S., Huang, C.-M.: Understanding User Reliance on AI in Assisted Decision-Making. Proceedings of the ACM on Human-Computer Interaction 6, 1-23 (2022)
53. Yoo, J.H., Kang, C., Lim, J.S., Wang, B., Choi, C.H., Hwang, H., Han, D.H., Kim, H., Cheon, H., Kim, J.W.: Development of an innovative approach using portable eye tracking to assist ADHD screening: a machine learning study. Front Psychiatry 15, 1337595 (2024)

How Work Feels: Affective Responses to Real-Effort Tasks

Akos Barta[1]; Eddie Begovic[1,] Victor Dinu[1,] Verena Dorner[1], Gerlinde Fellner-Röhling[1], Elisabeth Retegan[1], Luka Stefanovic[1]

[1] Vienna University of Economics and Business (WU), Vienna, Austria
{akos.barta, verena.dorner,gerlinde.fellner}@wu.ac.at
{eddie.begovic, victor.dinu, elisabeth.retegan, luka.stefanovic}@s.wu.ac.at

Abstract. Real-effort tasks are widely used in economic research to enhance external validity while maintaining experimental control. However, prior studies suggest individuals' affective and physiological responses to such tasks vary, challenging assumptions about their neutrality. This study partially replicates and extends a prior experiment by examining responses and performance across four tasks at two difficulty levels. In a lab experiment with 28 participants, we measured heart rate (HR) while participants completed math, ball, word finding and zero-counting tasks. We replicate evidence that task characteristics affect emotional engagement and potentially performance. Cluster analysis of HR patterns yields mixed results: two tasks show declining HR over time, while the other two vary substantially. This study helps develop a more nuanced view of the suitability of real-effort tasks for modelling real-world situations.

Keywords: heart rate • time series clustering • real-effort task

Introduction

Research on the interplay between decision-making, affective factors and decision task structure is attracting increasing attention across disciplines like NeuroIS [1], psychology [2], and economics [3], with the aim of identifying physiological mechanisms underlying behavior and choice [4, 5]. Many controlled experiments examine why decision-makers engage in effortful or costly actions [6, 7] with two experimental paradigms. In "stated effort", participants report intended effort by selecting from a range of values [8, 9]. In "real effort", participants exert cognitive or physical effort, such as solving math tasks [10] or throwing balls at a target [11]. A cost-effort function converts effort into a financial penalty that is subtracted from participants' compensation [12, 13]. Real-effort tasks aim to increase external validity by eliciting psychological and affective responses relevant in real-world decisions [3, 14].

But the real-effort paradigm has several limitations. First, a given task may evoke varying psychological responses across individuals [3, 15]. Second, different tasks may trigger distinct reactions within the same individual, potentially resulting in unpredictable behavior and uncontrolled variation in decision outcomes [3]. Affective reactions such as gratitude [16], anger [17], warm glow [18], and anxiety [15] have been proposed as explanations. Third, design-contingent factors like demand and ceiling effects may introduce bias into the observations [13].

While some research addresses the first [15, 19] and third issues [13], the second is largely unexplored. This paper partially replicates and extends a previous experiment on within-subject variations in reactions to real-effort tasks [20] and aims to contribute to answering the following research questions:

RQ1: *How do different real-effort tasks differ in their effects on subjective and physiological affective responses within the same participant?*

RQ2: *To what extent are task-related differences in subjective and physiological affective responses associated with differences in performance?*

Related Work and Hypotheses

Real-effort tasks. Real-effort tasks aim to elicit authentic behavioral and psychological reactions by requiring participants to exert creative, cognitive, or physical effort [3]. Tasks typically require solving puzzles, typing, doing calculations, doing physical clerical work, or small online games [13]. Research on affective reactions to such tasks is limited. Prior studies found that slider, math, and counting tasks induce higher reported anxiety than stated-effort tasks, though the slider task induces less anxiety than other real-effort tasks [15] and that arousal (measured as changes in HR) varies across participants in a slider task [19]. These studies compared between, not within, participants. One within-participants study found that HR arousal was lower in an engaging ball catching than a challenging math task, and that higher ball catching task enjoyment was correlated with higher valence and happiness scores [20].

Prior research in affective and emotional psychology suggests that individuals are likely to respond differently to real-effort tasks depending on the task characteristics and individuals' subjective experiences and perceptions of these tasks. Self-Determination Theory (SDT) proposes that tasks that support basic psychological needs (autonomy, competence, and relatedness) are associated with greater intrinsic motivation and more positive affective experiences [21]. Specifically, tasks that provide a sense of volition (autonomy), challenge and mastery (competence), and social connection (relatedness) tend to elicit greater interest, enjoyment, and satisfaction compared to tasks that are repetitive, externally controlled, or lack personal relevance [21, 22]. Apart from the objective task characteristics, perceived support of these needs changes individuals' affective reactions to tasks, which can lead to substantial inter-individual differences [21, 22].

Extending this view, Cognitive Appraisal Theory (CAT) posits that affective reactions are shaped by individuals' evaluations of events in relation to their goals, values, and situational constraints [23]. Tasks appraised as meaningful, engaging, or aligned with personal goals are more likely to elicit positive emotions such as interest or satisfaction. Similarly, Affective Events Theory (AET) proposes that work-related events, including task engagement, trigger affective reactions that vary across individuals depending on factors such as perceived challenge, reward, and personal relevance [24].

Taken together, SDT, CAT and AET strongly suggest that real-effort tasks are unlikely to be affectively neutral. By design, these tasks are intended to be repetitive and to have no intrinsic value so as not to distort effort provision within the controlled laboratory setting [3, 13]. Beliefs about performance in some tasks, such as math tasks, depend on competence or at least the individual's beliefs about their competence [3].

We expect affective responses to tasks to differ within individuals depending on how meaningful [23], engaging [22], challenging and repetitive [24] they are.

H1: Different real-effort tasks elicit different affective reactions in the same participant.

Based on SDT, we would expect that real-effort tasks which make individuals feel incompetent, rejected or oppressed give rise to psychological needs frustration, which in turn is associated with negative effect and negative performance outcomes [22]. Experimental evidence on the effect of affective reactions on performance in real-effort tasks is mixed. Anxiety was negatively associated with performance in a slider task [15], but another study found no link between affect and performance [19]. In the only within-subject study, higher task enjoyment increased effort in a math task but not a ball-catching task [20].

H2: Performance in real-effort tasks increases with positive affect.

Assessment of affective state. Affective states are typically assessed using self-report instruments [25].While they are simple to implement and inexpensive [26], they can be influenced by factors such as social desirability, self-esteem, and interoceptive abilities [26–28]. If the affective state needs to be assessed at several points in the experiment, administering a survey every time interrupts the experimental tasks and may alter participants' beliefs about the study as well as lead to participant fatigue and high time demands [28]. This makes biosignal-based approaches, which record physiological signals and infer affective states from changes in the signals [29], a good alternative for continuous and unobtrusive measurements not subject to self-report biases [29–31]. Measurement tools are now portable and affordable [30, 32]. HR-based measures, which are particularly easy to access [33], have been widely studied and correlate with changes in affective states [2, 29] like arousal [29] and stress [34, 35], although inferring discrete emotions from HR signals alone remains a challenge [1, 2, 36, 37].

Prior research suggests that HR and HRV reflect complex autonomic processes associated with emotional regulation rather than direct markers of specific emotional states [31, 37]. Variations in HR and HRV are understood to reflect regulatory capacity, flexibility, and the dynamic balance between sympathetic and parasympathetic activity, rather than discrete emotions per se [31, 37]. HR and HRV differ substantially between individuals, which complicates cross-subject comparisons. These differences reflect individual characteristics such as baseline autonomic tone, fitness, and regulatory capacity, as well as situational cues and different perceptions and reactions to these cues [31, 37, 38]. Similar HR and HRV patterns can be the result of different affective experiences, depending on contextual factors and individual differences [31, 37]. Conversely, different patterns may result from similar affective reactions. Another challenge, exacerbated in field settings, is the fact that HR signals are highly sensitive to methodological and contextual factors, including posture, respiration, and task conditions, which can introduce additional variability and make interpretation difficult [38].

Experimental Study Design

Treatments. We use a within-subject design with four real-effort tasks of different types (calculations, games, counting, puzzles), associated with different affective reactions. The math task, summing up integers [10], tends to induce anxiety [15, 39]. The ball-catching task, where participants catch randomly falling balls with a slider at the bottom of the screen, tends to induce engagement and enjoyment [40]. The grid search task, counting zeros in a grid [41], is considered very tedious [15]. The word finding task, finding anagrams from a series of letters, is considered to be a creative puzzle-solving type of task [42]. Since some literature suggests that the level of difficulty [11, 14] and the timing of the performance feedback [43, 44] play a role in shaping affective reactions, task difficulty was varied (high or low), and feedback was given either immediately after each task or delayed to the end of the experiment.

Procedure. Participants were equipped with Polar H10 HR Sensors, which provide highly reliable HR measurements [33, 45]. The experiment was implemented in oTree [46] and FRISBEE [47]. In *Part One*, participants received general instructions and a consent form, watched a 10-minute nature relaxation video [19, 48] and completed a short survey on their current affective and cognitive states (see below). The video served as a resting period to help participants achieve a relaxed state prior to starting the experiment, and to provide baseline HR measurement. In *Part Two*, participants played four rounds of tasks, in order of expected mental load to minimize fatigue effects (increasing mental load: ball task, counting zeroes, math, word finding). In each round, they received task instructions, played it for one minute, answered the survey on their current affective and cognitive state again and then watched 5-minute nature relaxation video as a washout period before proceeding to the next round. *Part Three* collected socio-demographic variables.

Measurement instruments (survey). Affective states are often measured with the Self-Assessment Manikin (SAM), which captures the dimensions of valence, arousal, and dominance through pictorial scales [49]. Changes in SAM are associated with changes of affective states like stress, sadness, and anxiety [50], and the SAM arousal dimension with changes in HR [49, 51]. The Discrete Emotions Questionnaire (DEQ) is another instrument with strong psychometric properties for assessing eight discrete emotional states: anger, disgust, fear, anxiety, sadness, desire, relaxation, and happiness [52]. The DEQ has been shown to reflect changes in emotional states after exposure to experimental stimuli [52]. The perceived cognitive state after each task was measured with three NASA-TLX subscales: perceived mental demand (of the task), perceived effort and perceived frustration [53].

Preliminary results

Participants. We recruited 28 students from a large European university (male=16, female=12). Sample size calculations for a repeated measures design, with medium effect size (f=0.25), significance level of at least 0.05, power of 0.8 and intrasubject correlation of at least 0.4 indicate that this is an appropriate sample size.

Task performance. Participants solved, on average, 7.71 math tasks (SD=4.9), caught 36.5 balls (SD=10.19), found 1.61 words (SD=1.031) and counted 2.82 zeroes (SD=2.97). When measuring performance on a percentage score (0-100%) and aggregating across tasks, participants performed better in easy than hard tasks (Mdn=70.63 and 52.61; U=165.5, $p<0.01$). High difficulty was perceived as more frustrating (U=37.5, $p<0.01$). Somewhat surprisingly, the word finding task was associated with the highest perceived effort and frustration, even exceeding the math task. As expected, the ball and counting tasks were associated with the lowest perceived mental load.

Hypothesis testing. Mixed effects regressions with random intercepts for every participant indicate that different tasks elicit different reactions (Table 1): HR and SAM valence are lower in the word counting task than the math task. Counting and word finding tasks made participants least happy. ***H1 is supported***. Increased difficulty was associated with lower valence, immediate feedback with lower perceived arousal.

Table 3. Regression Results

Predictor	HR	Valence	Arousal	Happiness	Performance
Task: Ball	*-1.62 (0.99)*	*0.50 (0.33)*	*0.00 (0.28)*	*0.54 (0.38)*	*-7.80 (7.28)*
Task: Count	*-1.28 (0.99)*	*-0.50 (0.33)*	*-0.46 (0.28)*	**-0.75 (0.38)***	**-23.25 (7.50)****
Task: Word	**-4.08 (0.99)*****	**-1.57 (0.33)*****	-0.04 (0.28)	**-1.54 (0.38)*****	**-48.73 (7.28)*****
Gender: Fem	*4.29 (4.19)*	*-0.21 (0.25)*	*-0.11 (0.24)*	*-0.06 (0.42)*	*4.57 (5.28)*
Difficulty: H	*-0.86 (4.15)*	**-0.69 (0.25)***	*0.03 (0.24)*	*-0.51 (0.42)*	**-17.85 (5.24)****
Feedback: I	*-6.66 (4.15)*	*0.15 (0.25)*	**-0.61 (0.24)***	*-0.20 (0.42)*	*4.66 (5.24)*
AIC	*699.41*	*386.90*	*357.84*	*436.42*	*1004.68*

*Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Random effects for participants*

Participants performed worst in counting and word finding tasks, and at high difficulty (Table 1, "Performance"). Higher levels of SAM valence are positively correlated with performance (β=5.75, SD=2.49, $p<0.05$) across all tasks and controlling for difficulty. ***H2 is supported.***

Overall model fit was strong (conditional R^2 ranging from 32.8% to 90.1%), except for arousal (17.2%). The fixed effects of task type and gender had moderate influence on subjective measures (26.3% for valence, 18.2% for happiness) but little impact on physiological responses (4.3% for heart rate). Our results suggest that individual differences between participants explain a large part of the variance in the changes in physiological measures.

Pairwise task contrasts show clear valence, happiness, HR and performance differences between certain tasks (Table 2). In line with our regression results, we see that participants perform worse in the word task than in the math task. In fact, participants perform worse in the word task than any other task. SAM valence for the word task is also lower than for any other task; in line with our expectations, SAM valence for the ball task is higher compared to the count task. Participants report greater happiness for the math and ball tasks compared to the word task, and for the ball task compared to the count task. HR is lower for the word task compared to all other tasks. All other contrasts are not significant.

Overall, we see surprisingly little differences between the math and ball tasks (none are significant) and the math and count tasks (except performance). The ball task is clearly enjoyed more than the count task, but this does not lead to a performance improvement. The word task is strictly dominated in terms of enjoyment and performance. SAM arousal does not differ between tasks.

Table 2. Pairwise task contrasts

	HR	Valence	Arousal	Happiness	Performance
math - ball	*1.618*	*-0.5*	*0*	*-0.536*	*7.802*
math - word	**4.083*** **	**1.571*** **	*0.036*	**1.536*** **	**48.731*** **
math - count	*1.283*	*0.5*	*0.464*	*0.75*	**22.385***
ball - word	**2.466***	**2.071*** **	*0.036*	**2.071*** **	**40.929*** **
ball - count	*-0.335*	**1.000***	*0.464*	**1.286** **	*14.583*
count - word	**2.801***	**1.071** **	*0.429*	*-0.786*	**26.346** **

*Note. * p < 0.05, ** p < 0.01, *** p < 0.001. Random effects for participants*

Clustering. We compare partitional clustering using Dynamic Time Warping (DTW) as the distance measure with Partitioning Around Medoids (PAM) for centroid calculation [54, 55], as used in [20], with k-shape clustering with shape-based distance (SBD), as used in [19], across k=2 to 4 clusters for our time-series analysis. We achieved the best silhouette width at 0.295 with DTW and k=2 (Figure 1), comparable to [19, 20]. Best overall accuracy of 39.3% was obtained with SBD and k=4, which is notably worse than the 66% reported in [20], but given that the current study compares 4 not 2 types of tasks, still better than the 25% given by chance.

HR patterns do not clearly vary with the four real-effort tasks (Figure 1), aside from a decline over time in the majority of ball and word task observations (cluster 1 includes 67% of ball and 75% of word task observations). In cluster 2, HR rises over time. Cluster 1 starts above baseline HR, cluster 2 slightly below. Math and count tasks are spread approximately evenly between the two clusters.

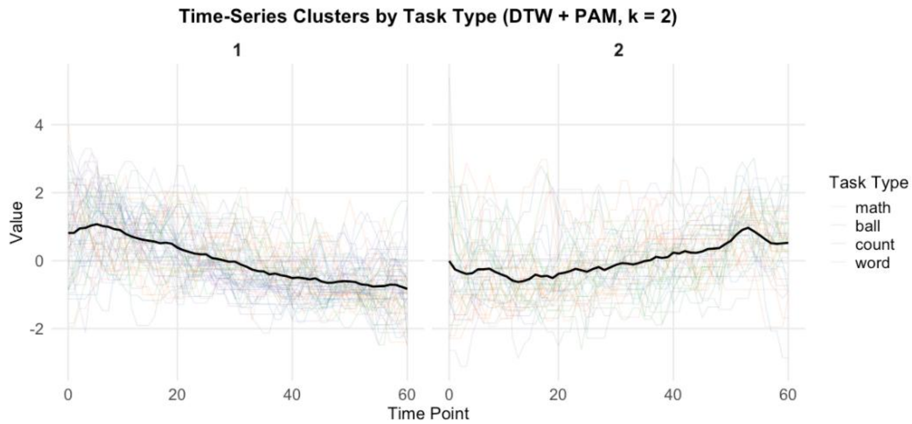


Fig. 9. Clustering Results.

Discussion and Future Research

This study partially replicates and extends a previous study [20] on the relationship between HR patterns, affective responses and performance in real-effort tasks, with the aim of improving experimental representations of real-world situations. In line with that study, we find differences in affective reactions: the word task is particularly likely to stress participants; the word and counting tasks are associated with least perceived happiness. Positive affect is associated with higher performance. Pairwise task comparisons show clearly that the word task is enjoyed less than any other task, and that performance in the word task is lower than in any other task. This warrants further research; our expectation was that creative tasks would be enjoyed more than math or count tasks, which are considered to induce anxiety and to be tedious, respectively [15, 39]. The classification of the word finding task as a creative task might be reconsidered, as far as that implies the assumption of greater enjoyment than other tasks. Future studies might use other real-effort tasks classified as creative, as well as "fun" tasks, to compare with the word finding task to establish whether there are systematic differences in enjoyment-related perceptions and performance.

HR patterns show a similar decline during the ball and word tasks. HR patterns in math and count tasks are mixed: these tasks were evenly spread between both clusters. Since task order was fixed, a general HR decline from ball to counting zeroes to math to word task might have been expected: without physical exertion, HR decreases over time. While we cannot exclude or quantify such an effect, our cluster analysis indicates that it was not the only effect: cluster 2 actually exhibits increasing HR on average and contains observations for tasks of all types. Cluster 1 exhibits declining HR and contains the majority of ball and word task observations. We conclude that, as frequently pointed out in literature, we observe large between-subject differences in HR patterns, which makes cross-subject comparisons difficult [38]. While our clustering approach clearly adds some value, as indicated by the better-than-chance classification, further research is needed to determine the limits of this approach.

Another challenge for this study and future studies on this topic remains the interpretation of HR patterns with respect to their affective correlates. More studies are required to determine whether, in the specific context of these and other real-effort tasks, such correlations between perceptions and physiological measures can be established. Subjective measures are inherently context-dependent, as participants anchor their responses to situational cues and internal reference frames [56], while HR measures are less clearly linked to context [37]. This is reflected in our results, where task type had a much larger effect on changes in perceived affect than changes in HR. By conducting similar studies with other populations and controlling for additional context- and personality-related factors, we aim to provide a clearer picture of the underlying processes yielding affective reactions to experimental tasks in the future.

Further research is needed to disentangle the effect of task types on arousal, as well as the effect of task parameters on affective reactions and performance. The connection between task type, mental load and affective state also warrants further investigation.

Limitations of this study are its small sample size and sociodemographic composition (students only). Future studies are needed to replicate these results and detect small effect sizes in other populations, as well as medium to small effect sizes when intrasubject correlation is low.

Overall, our results suggest that care is needed to choose the right experimental task such that the intended psychological reactions are elicited, and that no unintended affect or cognition is induced that could lead to systematic behavioral differences.

References

1. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS: Information Systems and the Brain. (2016).
2. Mauss, I.B., Robinson, M.D.: Measures of emotion: A review. Cognition and Emotion. 23, 209–237 (2009). https://doi.org/10.1080/02699930802204677.
3. Charness, G., Gneezy, U., Henderson, A.: Experimental methods: Measuring effort in economics experiments. Journal of Economic Behavior & Organization. 149, 74–87 (2018). https://doi.org/10.1016/j.jebo.2018.02.024.
4. Forte, G., Morelli, M., Casagrande, M.: Heart Rate Variability and Decision-Making: Autonomic Responses in Making Decisions. Brain Sciences. 11, 243 (2021). https://doi.org/10.3390/brainsci11020243.
5. Korosec-Serfaty, M., Riedl, R., Sénécal, S., Léger, P.-M.: 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 (2022). https://doi.org/10.3389/fnins.2022.883431.
6. Kurniawan, I.T., Guitart-Masip, M., Dolan, R.J.: Dopamine and effort-based decision making. Front Neurosci. 5, (2011).
7. Inzlicht, M., Shenhav, A., Olivola, C.Y.: The Effort Paradox: Effort Is Both Costly and Valued. Trends Cogn Sci. 22, 337–49 (2018).
8. Charness, G., Dufwenberg, M.: Promises and Partnership. Econometrica. 74, 1579–601 (2006).
9. Brandts, J., Ellman, M., Charness, G.: Let's Talk: How Communication Affects Con-tract Design. Journal of the European Economic Association. 14, 943–74 (2016).
10. Niederle, M., Vesterlund, L.: Do Women Shy Away From Competition? Do Men Compete Too Much? The Quarterly Journal of Economics. 122, 1067–1101 (2007). https://doi.org/10.1162/qjec.122.3.1067.
11. Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N.: Large stakes and big mistakes. Review of Economic Studies. 76, 451–469 (2009). https://doi.org/10.1111/j.1467-937X.2009.00534.x.
12. Smith, V.L.: Experimental Economics: Induced Value Theory. The American Economic Review. 66, 274–279 (1976).
13. Carpenter, J., Huet-Vaughn, E.: Real-effort tasks. In: Handbook of Research Methods and Applications in Experimental Economics. pp. 368–383. Edward Elgar Publishing (2019).
14. van Dijk, F., Sonnemans, J., van Winden, F.: Incentive systems in a real effort experiment. European Economic Review. 45, 187–214 (2001). https://doi.org/10.1016/S0014-2921(00)00056-8.

15. Lezzi, E., Fleming, P., Zizzo, D.J.: Does it Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize. SSRN Journal. (2015). https://doi.org/10.2139/ssrn.2594659.
16. Gneezy, U., List, J.A.: Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. Econometrica. 74, 1365–1384 (2006). https://doi.org/10.1111/j.1468-0262.2006.00707.x.
17. Gneezy, U., Imas, A.: Materazzi effect and the strategic use of anger in competitive interactions. Proceedings of the National Academy of Sciences. 111, 1334–1337 (2014). https://doi.org/10.1073/pnas.1313789111.
18. Brown, A.L., Meer, J., Williams, J.F.: Why Do People Volunteer? An Experimental Analysis of Preferences for Time Donations. Management Science. 65, 1455–1468 (2019). https://doi.org/10.1287/mnsc.2017.2951.
19. Greif-Winzrieth, A., Dorner, V., Könemann, J., Fellner-Röhling, G.: The Heart of Effort: Revealing Heart Rate Patterns in Real-Effort Tasks. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience. pp. 287–98. Springer Nature Switzerland, Cham (2025).
20. Dorner, V., Fellner-Röhling, G., Ferlini, G.A., Fink, N.S., Goldmann, J., Greif-Winzrieth, A., Schranzova, S., Tatar, C.-I., Tomek, R.: When Effort Feels Different: Physiological Evidence of Affective Variation in Real Effort Tasks. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., B. Randolph, A., and R. Müller-Putz, G. (eds.) Information Systems and Neuroscience. pp. 271–283. Springer Nature Switzerland, Cham (2025).
21. Deci, E.L., Ryan, R.M.: Self-Determination Theory. In: Handbook of Theories of Social Psychology: Volume 1. pp. 416–437. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom (2012). https://doi.org/10.4135/9781446249215.n21.
22. Thibault Landry, A., Zhang, Y., Papachristopoulos, K., Forest, J.: Applying self-determination theory to understand the motivational impact of cash rewards: New evidence from lab experiments. International Journal of Psychology. 55, 487–498 (2020). https://doi.org/10.1002/ijop.12612.
23. Schmidt, S., Tinti, C., Levine, L.J., Testa, S.: Appraisals, emotions and emotion regulation: An integrative approach. Motivation and Emotion. 34, 63–72 (2010). https://doi.org/10.1007/s11031-010-9155-z.
24. Weiss, H.M., Cropanzano, R.: Affective Events Theory: A Theoretical Discussion of the Structure, Causes and Consequences of Affective Experiences at Work. In: Staw, B.M. and Cummings, L.L. (eds.) Research in Organizational Behavior. pp. 1–74. JAI Press, Greenwich, CT (1996).
25. Dzedzickis, A., Kaklauskas, A., Bucinskas, V.: Human Emotion Recognition: Review of Sensors and Methods. Sensors. 20, 592 (2020). https://doi.org/10.3390/s20030592.
26. Robinson, M.D., Clore, G.L.: Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychological Bulletin. 128, 934–960 (2002). https://doi.org/10.1037/0033-2909.128.6.934.
27. Barrett, L.F., Quigley, K.S., Bliss-Moreau, E., Aronson, K.R.: Interoceptive Sensitivity and Self-Reports of Emotional Experience. Journal of Personality and Social Psychology. 87, 684–697 (2004). https://doi.org/10.1037/0022-3514.87.5.684.

28. Larsen, R.J., Fredrickson, B.L.: Measurement issues in emotion research. In: Well-being: The foundations of hedonic psychology. pp. 40–60. Russell Sage Foundation, New York, NY, US (1999).
29. Stangl, F.J., Riedl, R.: 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., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience [Internet. pp. 269–83. Springer International Publishing, Cham (2022).
30. Huhn, S., Axt, M., Gunga, H.-C., Maggioni, M.A., Munga, S., Obor, D.: The Im-pact of Wearable Technologies in Health Research: Scoping Review. JMIR mHealth and uHealth. 10:e34384, (2022).
31. Massaro, S., Pecchia, L.: Heart Rate Variability (HRV) Analysis: A Methodology for Organizational Neuroscience. Organizational Research Methods. 22, 354–393 (2019). https://doi.org/10.1177/1094428116681072.
32. Stangl, F., Riedl, R.: Measurement of Heart Rate and Heart Rate Variability with Wearable Devices: A Systematic Review. Wirtschaftsinformatik 2022 Proceedings. (2022).
33. Schaffarczyk, M., Rogers, B., Reer, R., Gronwald, T.: Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women. Sensors. 22, 6536 (2022). https://doi.org/10.3390/s22176536.
34. Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y.H., Koo, B.-H.: Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investig. 15, 235–245 (2018). https://doi.org/10.30773/pi.2017.08.17.
35. Fooken, J.: Heart rate variability indicates emotional value during pro-social economic laboratory decisions with large external validity. Sci Rep. 7, 44471 (2017). https://doi.org/10.1038/srep44471.
36. Siirtola, P., Tamminen, S., Chandra, G., Ihalapathirana, A., Röning, J.: Predicting Emotion with Biosignals: A Comparison of Classification and Regression Models for Estimating Valence and Arousal Level Using Wearable Sensors. Sensors. 23, 1598 (2023). https://doi.org/10.3390/s23031598.
37. Appelhans, B.M., Luecken, L.J.: Heart Rate Variability as an Index of Regulated Emotional Responding. Review of General Psychology. 10, 229–240 (2006). https://doi.org/10.1037/1089-2680.10.3.229.
38. Laborde, S., Mosley, E., Thayer, J.F.: Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting. Front. Psychol. 08, (2017). https://doi.org/10.3389/fpsyg.2017.00213.
39. Shostak, B.B., Peterson, R.A.: Effects of anxiety sensitivity on emotional response to a stress task. Behaviour Research and Therapy. 28, 513–21 (1990).
40. Gächter, S., Huang, L., Sefton, M.: Combining "real effort" with induced effort costs: the ball-catching task. Experimental Economics. 19, 687–712 (2016).
41. Abeler, J., Falk, A., Goette, L., Huffman, D.: Reference Points and Effort Provision. American Economic Review. 101, 470–492 (2011). https://doi.org/10.1257/aer.101.2.470.
42. Charness, G., Villeval, M.-C.: Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory. American Economic Review. 99, 956–978 (2009). https://doi.org/10.1257/aer.99.3.956.

43. Brehm, J.W., Self, E.A.: The Intensity of Motivation. Annu. Rev. Psychol. 40, 109–131 (1989). https://doi.org/10.1146/annurev.ps.40.020189.000545.
44. Richter, M., Gendolla, G.H.E., Wright, R.A.: Three Decades of Research on Motivational Intensity Theory. In: Advances in Motivation Science. pp. 149–186. Elsevier (2016). https://doi.org/10.1016/bs.adms.2016.02.001.
45. Vermunicht, P., Makayed, K., Meysman, P., Laukens, K., Knaepen, L., Vervoort, Y.: Validation of Polar H10 chest strap and Fitbit Inspire 2 tracker for measuring continuous heart rate in cardiac patients: impact of artefact removal algorithm. Euro-pace. 122, (2023).
46. Chen, D.L., Schonger, M., Wickens, C.: oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance. 9, 88–97 (2016). https://doi.org/10.1016/j.jbef.2015.12.001.
47. Gorny, P.M., Greif-Winzrieth, A., Nieken, P., Wuest, F.A., FrISBEE, B.N.: Frame-work for integrating sensor data in behavioral and economic experiments.
48. Piferi, R.L., Kline, K.A., Younger, J., Lawler, K.A.: An alternative approach for achieving cardiovascular baseline: viewing an aquatic video. International Journal of Psychophysiology. 37, 207–217 (2000). https://doi.org/10.1016/S0167-8760(00)00102-1.
49. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry. 25, 49–59 (1994). https://doi.org/10.1016/0005-7916(94)90063-9.
50. Houben, M., Den Noortgate, W., Kuppens, P.: The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. Psychological Bulletin. 141, 901–30 (2015).
51. Bradley, M.M., Lang, P.J.: Emotion and Motivation. In: Cacioppo, J.T., Tassinary, L.G., and Berntson, G. (eds.) Handbook of Psychophysiology [Internet. pp. 581–607. Cambridge University Press, Cam-bridge (2007).
52. Harmon-Jones, C., Bastian, B., Harmon-Jones, E.: The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-Reported Emotions. PLOS ONE. 11, e0159915 (2016). https://doi.org/10.1371/journal.pone.0159915.
53. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A. and Meshkati, N. (eds.) Human Mental Workload. pp. 139–183. North-Holland (1988). https://doi.org/10.1016/S0166-4115(08)62386-9.
54. Paparrizos, J., Gravano, L.: k-Shape: Efficient and Accurate Clustering of Time Series. SIGMOD Rec. 45, 69–76 (2016). https://doi.org/10.1145/2949741.2949758.
55. Sardá-Espinosa, A.: Time-Series Clustering in R Using the dtwclust Package. The R Journal. 11, 22 (2019). https://doi.org/10.32614/RJ-2019-023.
56. Teoh, Y.Y., Cunningham, W.A., Hutcherson, C.A.: Framing Subjective Emotion Reports as Dynamic Affective Decisions. Affec Sci. 4, 522–528 (2023). https://doi.org/10.1007/s42761-023-00197-y.

“I Don’t Really Care”: How Generative AI Use Shapes Affective Reactions to Feedback on Work Outcomes

Lucas Memmert, Eva Bittner

University of Zurich, Zurich, Switzerland
{lucas.memmert,eva.bittner}@uzh.ch

Abstract. Knowledge workers increasingly use generative artificial intelligence (GenAI) to co-create work outcomes. While feedback on work can evoke affective reactions that shape subsequent behavior, it remains unclear how GenAI-assisted work influences these reactions. Drawing on affective events theory and research on effort investment, we examine how using GenAI shapes individuals’ affective reactions to feedback on their work outcomes and the extent to which they revise them. We conduct a controlled between-subject experiment in which participants complete writing tasks either with or without GenAI. After each task, participants receive standardized feedback while neurophysiological reactions (feedback-related negativity, FRN) are recorded using EEG. We propose that GenAI use reduces effort investment, leading to weaker affective reactions to feedback (reflected in attenuated FRN amplitude irrespective of feedback valence) and greater extent of revision for the work outcomes. The study contributes to research on consequences on human–AI collaboration for individuals and feedback reactions.

Keywords: Affective reactions · Feedback · GenAI · Effort · Feedback-related negativity

Introduction

Knowledge workers increasingly use generative artificial intelligence systems (GenAI) to create work outcomes [1]. Unlike prior tools, GenAI can contribute directly to those outcomes, leading to what can be referred to as the co-creation with AI systems [2]. In many cases, these work outcomes are shared with colleagues, superiors, or clients, who then provide feedback; in fact, feedback is an essential aspect of knowledge work in organizations [3].

Literature shows that people can experience affective reactions when they receive feedback on their work [4, 5]. Affective reactions are dependent on several factors, such as personal investment into the work outcomes, reflected, for example, by time or effort spent, as well as the type of feedback provided [6]. Affective reactions to feedback can then influence subsequent behavior [4], influencing, for example, reluctance to revise work results, but can also influence performance on unrelated tasks [5].

Prior literature shows that working with GenAI alters work processes and, consequently, the relationship between workers and their work outcomes [7]. With workers having to spend less effort to achieve a work outcome [7, 8], they may also be less personally invested, show weaker, that is, less intense, affective reactions to feedback, and be more open to revising a work outcome. This could fundamentally change the way of working in knowledge work, making it more iterative and fostering co-creation within teams, as people are less protective of their own initial ideas and more willing to revise. In this study, we therefore explore the following research question: *How does working with GenAI influence individuals' affective reactions to feedback on their work outcomes and their extent of revision of those work outcomes?*

To investigate this question, we conduct a between-subject lab experiment. Participants complete a writing task without or with GenAI assistance. After completing the task, participants receive standardized, random feedback on their work. Affective reactions to feedback are measured neurophysiologically (EEG) to determine whether working with GenAI attenuates these reactions. We then ask participants to revise their text and measure revision behavior.

Working with GenAI fundamentally changes how knowledge work is conducted, and how workers feel about the results they produce [7]. With this study, we seek to clarify how working with GenAI shapes affective reactions to feedback. We contribute to the emerging human-AI collaboration literature [9], specifically exploring the dynamics in human-AI dyads [10], as well as to the literature on affective reactions to feedback [11]. This study will have practical implications, providing insights into how GenAI-assisted work may influence feedback reactions and iteration behavior in knowledge work.

Background and Hypotheses Development

Affective Reactions to Feedback on Work Outcomes

Feedback on one's work is an essential practice in organizations [3]. Both positive and negative feedback can trigger affective reactions [5], particularly when they are surprising, deviating from expectations [12]. According to affective events theory, workplace events such as evaluations or criticism evoke affective reactions that shape subsequent attitudes and behaviors [13]. These affective reactions influence how individuals respond to feedback, including their defensiveness toward criticism and their willingness to revise their work. The intensity of affective reactions depends on how personally invested or involved individuals feel in the evaluated outcome, or the extent to which the outcome is contingent on their actions [5, 14, 15].

The neurological reaction to feedback can be measured via event-related potentials (ERPs) in electroencephalographic (EEG) signals [16]. One such component, the feedback-related negativity (FRN), reflects early evaluative processing of feedback outcomes. FRN has traditionally been interpreted as being sensitive to feedback valence [6], however, it may reflect the magnitude of deviation from expectations, rather than whether outcomes are positive or negative per se; so larger amplitudes may be observed for outcomes that differ more strongly from what was anticipated, i.e., reflecting surprise [16].

Importantly, FRN amplitude is not only determined by expectation deviations but is also influenced by the degree to which outcomes are perceived as contingent on one's own actions. Specifically, FRN amplitudes are reduced when individuals merely observe outcomes rather than actively generate them, and this reduction is associated with lower reported subjective involvement [15]. Thus, the affective reactions to feedback and the neural response may depend not only on expectation deviations but also on how the outcome was produced.

Generative AI, Effort Investment, and Feedback Reactions

GenAI is increasingly used by knowledge workers to create work outcomes [1, 17, 18], with the GenAI directly contributing (sometimes significant portions) to work results. Recent literature suggests that compared to human-only, unassisted work, GenAI-assisted work can influence how knowledge workers feel about their work results, including diffused felt responsibility or ownership [8, 19–21]. Moreover, when GenAI takes over certain aspects of the task, it may reduce the (felt) effort required to produce work outputs [7, 8, 22]. This is relevant, because effort investment and how much an outcome is valued are associated [7, 14]. Hence, if GenAI reduces one's effort investment for achieving a task outcome, that outcome may feel less tied to one's actions and less personally significant or meaningful. Aligned with this suggestion, Kosmyna et al. find that "essays written with the help of LLM carried a lesser significance or value to the participants" [21]. Moreover, they show that working with GenAI as compared to working alone leads to a difference in neural activity [21]. However, other studies show that GenAI use may lead to additional effort, for example, for prompting or fact checking [8, 23], which is why we will measure felt effort as part of our experiment.

Based on the above, we suggest that evaluative feedback on the work outcome may elicit a weaker affective reaction when the work result was produced with a GenAI. More specifically, we expect that working with GenAI is associated with a reduced FRN amplitude. Importantly, FRN has been shown to reflect deviations from expectations irrespective of feedback valence [16]. Accordingly, the proposed dampening effect is expected to occur for both positive and negative feedback. While in this study we focus on effort as a mediating construct, we call for future research to explore additional pathways on how GenAI shapes reactions to feedback.

Hypotheses

Accordingly, we propose the following conceptual model for our study (Fig. 1), with the three hypotheses below: a direct effect of working with a GenAI on the affective reaction to feedback (H1), partial mediation through effort (H2), and a behavioral consequence in the effect on the extent of revision for work outcomes (H3).

We do not theorize full mediation of affective reaction to feedback through effort (H2), because, as discussed above, GenAI use may influence worker behavior and perception in a variety of ways including perceived ownership or responsibility [8]. Because affective reaction to feedback is known to influence subsequent behavior, we theorize it influences revision behavior, that is, the extent of revision. However, we acknowledge that other explanations may be possible, for example, working with GenAI may affect individuals' confidence in the generated output or their perceived authorship.

Hypothesis 1: Working with GenAI directly reduces individuals' affective reaction to feedback on their work outcomes.

Hypothesis 2: The relationship between GenAI use and affective reaction to feedback is partially mediated by effort, so that working with GenAI reduces effort (H2a), and reduced effort reduces the affective reaction to feedback (H2b).

Hypothesis 3: Lower affective reaction to feedback increases extent of revision of work outcomes.

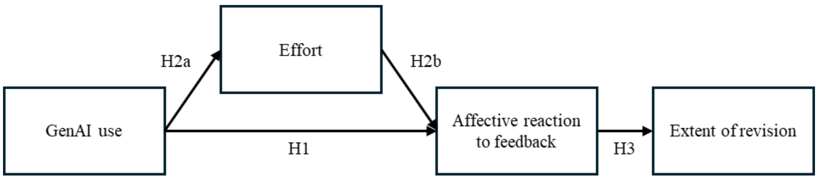


Fig. 1. Conceptual model

Method

Design

To investigate how working with GenAI influences affective reactions to feedback on work outcomes, the study uses a between-subject experimental design with two conditions: (1) *human-only*, in which participants complete a series of four writing tasks within their domain knowledge without GenAI assistance, and (2) *human+GenAI*, in which participants complete the same tasks with the assistance of a GenAI, such as ChatGPT (see [7]). The emotional reactions are measured neurophysiologically directly after feedback presentation.

Participants

Participants are recruited from a university student pool. Students are appropriate participants because they are familiar with writing tasks and the topic, and they themselves become future knowledge workers. Participants receive course credit for participation. Participants will be split equally across conditions.

Procedure

The study follows a pre-defined procedure (Fig. 2) as described below. In the *Introduction and pre-questionnaire*, participants first provide informed consent, and fill in a pre-questionnaire on demographics, prior GenAI experience, and baseline mood. They are informed that they will write several short essays and receive immediate feedback on those essays along several dimensions. They will be informed that the feedback will be based on an automatic evaluation by a specialized AI system trained on a database of prior submissions. To increase feedback credibility, the AI is described as having shown high agreement with human expert raters. Moreover, a brief delay prior to the AI giving feedback to the participants will be added.

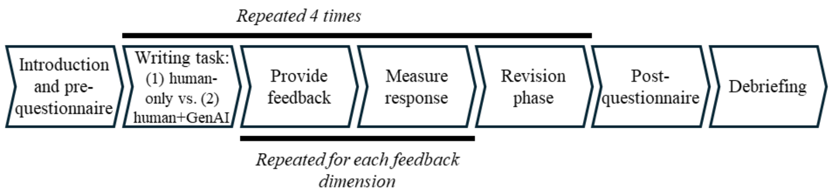


Fig. 2. Experiment procedure

As part of the *writing task*, participants write a short essay of about 250 words within a 4-minute time limit. In the human-only condition, participants write without GenAI support; in the human+GenAI condition, participants are encouraged to use a GenAI tool (ChatGPT) to assist with writing.

After completing each essay, participants receive multidimensional *feedback* on their text, for example, on its structure or the quality of its arguments. The feedback is presented sequentially to allow for measuring emotional reactions to each feedback statement. For each dimension, the feedback states whether performance is above average or below average compared to prior submissions of other participants with respect to the feedback dimension in the form:

"Compared to submissions from other participants, the <feedback dimension, e.g., clarity of your argument> is <feedback direction, e.g., below average>."

To ensure comparable exposure to feedback events across participants, feedback is randomized rather than based on actual essay quality (also see [11]). This allows us to isolate the effects of the experimental manipulation and avoid performance-related confounds, because essays produced alone or with GenAI support may systematically differ across quality dimensions. We acknowledge that this may reduce feedback credibility in some cases. Therefore, we will assess and control for three related but distinct aspects: participants' perceived surprise of each feedback event as a direct indicator of subjective expectation deviation (reflecting a potential "standard" participants may compare the feedback against; see [5]), post-experimental feedback credibility or suspicion, and the objective fit between assigned feedback and independently scored essay quality as a proxy for feedback plausibility. This allows us to control for to what extent the random feedback reflected actual quality, and to exclude suspicious participants. Participants will receive equal amounts of positive and negative feedback.

After receiving feedback on all dimensions, participants are asked to revise their text for 2 minutes, capturing their extend of revision of the work outcome. Importantly, none of the participants will have access to GenAI in this step. This prevents participants in the GenAI condition to merely re-generate the essay according to the feedback. The steps from writing to revision will be repeated four times per participant with different essay topics. The order of topics will be randomized. After completing the revision for the last essay, participants will complete a post-questionnaire (see next section). In the *debriefing*, participants are informed that the feedback was randomized and not based on actual text, and that this was necessary to ensure standardized experimental conditions.

Measurement

The between-subject experiment uses a binary manipulation: (1) human-only, that is, writing without GenAI support, vs. (2) human+GenAI, that is, writing with GenAI support. Participants will be randomly assigned to either condition and will remain in that condition for all four writing tasks.

Affective reaction to feedback is our central dependent variable. As discussed in Section 2, we operationalize affective reactions to feedback using the FRN, an ERP associated with early evaluative processing of feedback outcomes. FRN is a negative deflection occurring approximately 200–300ms after feedback onset [16]. Prior research shows that FRN amplitude reflects the evaluation of outcomes relative to expectations, with larger (i.e., more negative) amplitudes observed for outcomes that are unexpected or carry greater affective significance [16]. In line with this interpretation, we use FRN amplitude as a neural indicator of the intensity of evaluative processing of feedback. Reduced affective reactions are expected to manifest in attenuated FRN amplitudes (i.e., less negative deflections). FRN will be analyzed as the maximal amplitude within a predefined time window [16, 24] 200–300 ms following feedback onset at the frontal midline electrode (Fz, [6, 24]), relative to a pre-feedback baseline. Because physiological measurements can be noisy, multiple feedback events are used to obtain reliable estimates [24], i.e., one feedback event and reaction measurement per feedback dimension per essay topic.

Following Campbell et al. [7], to measure effort, we combine self-reported perceived task effort using the NASA-TLX task load scale [25] and behavioral proxies such as the number of characters typed. To measure the revision extent, we combine measures like the number of edits made, the number of characters changed, and the textual similarity between the original and revised texts using semantic embeddings.

Moreover, we will collect several control variables, including perceived credibility of the feedback (allowing us to exclude participants overly suspicious of the feedback), perceived autonomy during task completion, attitude towards AI, and baseline mood [13] as well as self-relevance of output. Additionally, we will calculate a feedback-fit score, relating essay quality (automated scoring via LLM, see [7]) to the standardized, random feedback the participant received, allowing us to control for the reasonably expected feedback-performance-gap.

Ethical Considerations

The study involves partial deception, as feedback is randomized rather than based on actual performance. Therefore, the participants are fully debriefed after the experiment. The study is reviewed by the institutional review board (ethics committee).

Expected Contribution

This study contributes to the emerging discourse on human-GenAI work, shedding light on human-AI dynamics [10] and how working with GenAI affects workers [8]. More broadly, it advances our understanding of IS use by complementing self-reports through neurophysiological measures as suggested by Dimoka et al [26]. More specifically, the study identifies and tests a mechanism through which GenAI use shapes reactions to feedback. We propose that working with GenAI reduces effort investment in producing work outcomes, which in turn attenuates affective reactivity to evaluative feedback, as reflected in reduced FRN amplitude. By linking effort investment, neurophysiological reactions to feedback, and subsequent revision behavior, the study provides a process-level account of how human–AI collaboration alters feedback processing.

Providing and receiving feedback is essential within organizations [3, 5]. From a practical perspective, this mechanism suggests that GenAI-assisted work may reduce resistance to feedback and thereby foster more iterative work practices. As a result, individuals may be more open to revise intermediate outputs and engage in improvement cycles, which has implications for the feedback culture in GenAI-supported work environments.

References

1. Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R.: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School Working Paper, vol. (2023)
2. Banh, L., Strobel, G.: Generative artificial intelligence. Electron Markets, vol. 33 (2023). doi: 10.1007/s12525-023-00680-1
3. Ilgen, D.R., Fisher, C.D., Taylor, M.S.: Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, vol. 64, 349–371 (1979). doi: 10.1037/0021-9010.64.4.349
4. Hackman, J., Oldham, G.R.: Motivation through the design of work: test of a theory. Organizational Behavior and Human Performance, vol. 16, 250–279 (1976). doi: 10.1016/0030-5073(76)90016-7
5. Kluger, A.N., DeNisi, A.: The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, vol. 119, 254–284 (1996). doi: 10.1037/0033-2909.119.2.254
6. Huang, Y., Yu, R.: The feedback-related negativity reflects "more or less" prediction error in appetitive and aversive conditions. Frontiers in neuroscience, vol. 8, 108 (2014). doi: 10.3389/fnins.2014.00108

7. Campbell, A.V., Wang, Y., Inzlicht, M.: Experimental evidence that exerting effort increases meaning. Cognition, vol. 257, 106065 (2025). doi: 10.1016/j.cognition.2025.106065
8. Memmert, L., Soroko, D., Bittner, E.: From Effort Reduction to Effort Management: An Expectancy Theory Perspective on Professionals' Work Practices with Generative AI. Bus Inf Syst Eng, vol. , 615–635 (2025). doi: 10.1007/s12599-025-00960-4
9. Riemer, K., Peter, S., Schwabe, G., Chatterjee, S., Adam, M., Davison, R.M.: Generative AI Is Neither Just Another IT Artefact nor a Colleague: Methodological Guidance for IS Scholarship. Information Systems Journal, vol. (2025). doi: 10.1111/isj.70027
10. Makarius, E.E., Mukherjee, D., Fox, J.D., Fox, A.K.: Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. Journal of Business Research, vol. 120, 262–273 (2020). doi: 10.1016/j.jbusres.2020.07.045
11. Ilies, R., Pater, I.E. de, Judge, T.: Differential affective reactions to negative and positive feedback, and the role of self-esteem. Journal of Managerial Psychology, vol. 22, 590–609 (2007). doi: 10.1108/02683940710778459
12. Kluger, A.N., Lewinsohn, S., Aiello, J.R.: The Influence of Feedback on Mood: Linear Effects on Pleasantness and Curvilinear Effects on Arousal. Organizational Behavior and Human Decision Processes, vol. 60, 276–299 (1994). doi: 10.1006/obhd.1994.1084
13. Weiss, H.M., Cropanzano, R.: Affective events theory. Research in Organizational Behavior, vol. 18, 1–74 (1996)
14. Norton, M.I., Mochon, D., Ariely, D.: The IKEA effect: When labor leads to love. J Consum Psychol, vol. 22, 453–460 (2012). doi: 10.1016/j.jcps.2011.08.002
15. Yeung, N., Holroyd, C.B., Cohen, J.D.: ERP correlates of feedback and reward processing in the presence and absence of response choice. Cerebral cortex (New York, N.Y. : 1991), vol. 15, 535–544 (2005). doi: 10.1093/cercor/bhh153
16. Hauser, T.U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., Brem, S.: The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. NeuroImage, vol. 84, 159–168 (2014). doi: 10.1016/j.neuroimage.2013.08.028
17. Edmonds, L.: 3 years old, 800 million users, 29,000 prompts a second: ChatGPT's meteoric rise, by the numbers. Business Insider, vol. (2025)
18. Benbya, H., Strich, F., Tamm, T.: Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. Journal of the Association for Information Systems, vol. 25, 23–36 (2024). doi: 10.17705/1jais.00861
19. Memmert, L., Tavanapour, N.: Towards Human-AI Collaboration in Brainstorming: Empirical Insights into the Perception of Working with a Generative AI. BRAINSTORMING-QUAL. In: Proceedings of the 31st European Conference on Information Systems (ECIS) (2023)
20. Stieglitz, S., Mirbabaie, M., Möllmann, N.R.J., Rzyski, J.: Collaborating with Virtual Assistants in Organizations: Analyzing Social Loafing Tendencies and Responsibility Attribution. Information systems frontiers, vol. 24, 745–770 (2022). doi: 10.1007/s10796-021-10201-0
21. Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.-H., Beresnitzky, A.V., Braunstein, I., Maes, P.: Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv (2025)

22. Memmert, L., Cvetkovic, I., Tavanapour, N., Bittner, E.: Brainstorming with a Generative Language Model: Effect of Exposure to AI Ideas on Brainstorming Performance and Cognitive Load. Bus Inf Syst Eng, vol. (2025). doi: 10.1007/s12599-025-00974-y
23. Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A.E., Sarkar, A., Sellen, A., Rintel, S.: The Metacognitive Demands and Opportunities of Generative AI. In: Mueller, F.F., Kyburz, P., Williamson, J.R., Sas, C., Wilson, M.L., Dugas, P.T., Shklovski, I. (eds.) Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–24. ACM, New York, NY, USA (2024). doi: 10.1145/3613904.3642902
24. Marco-Pallares, J., Cucurell, D., Münte, T.F., Strien, N., Rodriguez-Fornells, A.: On the number of trials needed for a stable feedback-related negativity. Psychophysiology, vol. 48, 852–860 (2011). doi: 10.1111/j.1469-8986.2010.01152.x
25. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research Human Mental Workload. Advances in Psychology, vol. 52, pp. 139–183. Elsevier (1988). doi: 10.1016/S0166-4115(08)62386-9
26. 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., et al.: On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for Neurois1. MIS Quarterly, vol. 36, 679–702 (2012). doi: 10.2307/41703475

Generative Artificial Intelligence–Enabled Cognitive Offloading: Effects on Task Performance and Skill Development

Philipp Pecher[1,2] ,and Aycan Aslan[3]

[1] University of Goettingen, Goettingen, Germany
philipp.pecher@uni-goettingen.de
[2] Macquarie University, Sydney, Australia
philipp.pecher@students.mq.edu.au
[3] IT University Copenhagen, Copenhagen, Denmark
ayas@itu.dk

Abstract. This research-in-progress investigates cognitive offloading as a key mechanism through which generative AI (GenAI) use influences task performance and skill development. Drawing on cognitive offloading literature, we conceptualize distinct offloading modalities that differentially shift necessary or unnecessary mental load to GenAI systems. We argue that these offloading modalities have opposing effects on skill development, either supporting or inhibiting it. To examine these mechanisms, we propose a neurophysiological experiment in which participants complete a software development task with GenAI support. Using electroencephalography (EEG), we measure differences in mental load associated with alternative offloading modalities and relate them to both task performance and subsequent skill development. By directly assessing cognitive load during GenAI use, this study seeks to uncover the cognitive processes underlying the ambivalent effects of GenAI on task performance and skill development, thereby advancing the theoretical understanding of cognitive offloading in GenAI-enabled knowledge work.

Keywords: Cognitive offloading · skill development · task performance · GenAI · deskilling · upskilling

Introduction

With the diffusion of generative AI (GenAI) into organizational life, these systems are becoming increasingly relevant for knowledge-intensive work [1] and have already been shown to improve individual task performance [2–5]. Examples include the growing use of tools such as GitHub Copilot to support software development [2], the use as an AI assistant for customer support agents [3], and the use of GenAI to generate and refine advertising slogans [4].

In this context, emerging research indicates that GenAI use shapes individuals' skill development in *ambivalent* ways [3, 6]. On one hand, studies indicate that the use of GenAI positively affects skill development, for example in the context of language skills [3]. On the other hand, first studies also show that GenAI use might have a negative effect on skills due to overreliance, for example in the context of software development [6].

Accordingly, it remains important to reveal the *underlying mechanisms* driving how GenAI use influences skill development. While early work has demonstrated direct effects of GenAI use on skills [3, 6], it has paid less attention to the cognitive mechanisms that could account for these effects, leaving the "why" and "how" insufficiently explained.

To fill this gap, we draw on cognitive science literature, focusing on *cognitive offloading* [7]. Cognitive offloading refers to relying on external actions or artifacts to reconfigure a task's information-processing requirements and thereby reduce cognitive effort [8]. In the context of GenAI-supported knowledge work, we argue that this perspective offers a useful lens for explaining the cognitive mechanisms through which GenAI use shapes skill development. Moreover, we will examine how cognitive offloading relates to task performance, allowing us to identify potential differences between its effects on skill development and its effects on performance. We therefore ask: *How does GenAI-enabled cognitive offloading, affect users' skill development and task performance?*

To address our research question, we propose a neurophysiological experiment using electroencephalography (EEG) [9] to examine GenAI-enabled cognitive offloading in a software development task. While prior studies primarily document high-level effects of GenAI use on task performance and skill development [3, 6], our approach is well suited to uncover the *underlying cognitive mechanisms* that may account for these effects [10, 11]. This study aims to contribute to the literature by challenging established assumptions about the interrelation between task performance and skill development in the context of cognitive offloading. We expect that the unstructured use of GenAI will disrupt previously observed patterns in which improvements in task performance are aligned with skill development [12, 13].

Background

The Interdependence of Skill and Task Performance

Skill can be conceptualized as the integration of declarative and procedural knowledge [14]. Declarative knowledge refers to knowledge about things and facts [15, 16], that is, what an individual knows. Procedural knowledge, in contrast, captures tacit knowledge namely, knowledge of how to carry out a task. Knowing that a recipe requires a specific sequence of ingredients is an example of declarative knowledge, whereas being able to prepare the dish by following and executing the cooking steps is an example of procedural knowledge.

Task performance and skill development are interrelated, as learning-by-doing leads to the continuous accumulation of problem-solving experience [13], which in turn directly results in a significant reduction in average task completion times and an increase in productivity [12]. Conversely, performing a range of tasks enhances skill development by helping employees integrate abstract principles and latent schemas, which in turn improves their efficiency in tackling new or different types of problems [12].

Prior research on skill development demonstrates that both the format and the timing of information delivery significantly shape the development of skills and task performance [15, 17, 18]. Generative AI introduces a novel distinction relative to prior information technology artifacts in that it is capable of performing complex tasks autonomously, rather than merely supporting or structuring human information processing [19, 20]. This expanded capability fundamentally alters how information is delivered and processed during task execution, with important implications for the relationship between task performance and skill development [21].

Cognitive Offloading

When someone writes information down to remember it later, they engage in cognitive offloading, defined as "the use of physical action to alter the information processing requirements of a task so as to reduce cognitive demand" ([8], p. 677). Traditional examples include using a calculator to solve a math problem or relying on a shopping list to remember items in the supermarket [8]. GenAI provides substantial expansion opportunities by enabling users to offload not only fragments of information but potentially broader portions of tasks and their execution to a single system through human-like conversational interaction which reduces the effort of cognitive offloading substantially [22].

Previous research shows that cognitive offloading is influenced by both task characteristics and individual cognitive and metacognitive factors. Offloading behavior increases when memory load is high making external aids more attractive [23]. In time-critical contexts, individuals often engage in cognitive offloading as these actions can be faster and more reliable than internal computations alone [22]. Beyond objective task demands, subjective self-confidence plays a crucial role. Lower self-confidence reliably predicts greater reliance on memory aids, even when self-confidence is experimentally decoupled from actual performance [24]. Consistent with this, the skill level does not deterministically predict cognitive offloading. Instead, metacognitive evaluations explain additional variance, and cognitive offloading has been shown to benefit individuals across a wide range of ability levels [23]. Finally, cognitive offloading is strongly influenced by the effort required to do so. Increased physical or procedural effort (e.g. additional steps or access restrictions) reduces cognitive offloading [22]. Conversely, when the effort required for cognitive offloading is minimized, as in the case of GenAI systems, a larger portion of the mental load is likely to be shifted to the systems.

Hypothesis Development

Cognitive offloading acts as a crucial mechanism for improving task performance by providing environmental support that reduces internal cognitive demands, thereby enabling the reallocation of resources for the execution tasks [25, 26]. Nevertheless, reliance on external memory comes with costs [6, 14], including reduced depth of processing and a "directed forgetting" effect that occurs when individuals expect continuous access to outsourced information [25, 26].

We conceptualizes two modes of cognitive offloading in the context of skill development. *Substitutive offloading* refers to the condition in which GenAI replaces mental processes that are essential for skill development, thereby offloading both unnecessary and necessary mental load [10, 11]. Although this form of offloading may enhance task performance by reducing immediate cognitive demands, it is expected to undermine skill development by bypassing the cognitive effort required for internalization and competence formation [26, 27].

In contrast, *constructive offloading* describes the use of external support to reduce only unnecessary mental load, such as routine or peripheral processing demands, while maintaining engagement in the core cognitive activities required for learning [10, 11]. By scaffolding rather than replacing essential mental processing, *constructive offloading* is expected to support task performance while simultaneously fostering skill development over time [28]. Constructive offloading reduces only unnecessary mental load, meaning cognitive burden that does not contribute to skill development, while preserving the mental load essential for learning. Consequently, users in the constructive offloading condition outperform those in the no-offloading condition on task performance, yet underperform relative to those in the substitutive offloading condition. Critically, however, because substitutive offloading eliminates all mental load, including that which is essential for skill acquisition, the constructive offloading condition yields superior skill development compared to both the no-offloading and substitutive offloading conditions. Thus, we hypothesize:

Hypothesis 1a. Task performance will be higher in the *constructive offloading* condition than in the *no offloading* condition.

Hypthesis 1b. Task performance will be higher in the *substitutive offloading* condition than in the *constructive offloading* condition.

Hypothesis 2a. Skill development will be higher in the *no offloading* condition than in the *substitutive offloading* condition.

Hypothesis 2b. Skill development will be higher in the *constructive offloading* condition than in the *no offloading* condition.

Together, these hypotheses formalize a central tension in GenAI-enabled work: *substitutive offloading* reaches hightest task performance but can undermine the development of skills , whereas *constrctuve offloading* reaches highest level of skill development on a lower level of task performance.

Proposed Research Design

Experimental Design. We will conduct a neurophysiological laboratory experiment to examine how GenAI-enabled cognitive offloading, implemented through different cognitive offloading modalities, affects task performance and skill development in a software development context. To capture cognitive offloading, we will use EEG-based measures which are widely used in research on information systems [9, 29, 30]. We aim to recruit N = 60 participants, the majority of which will be information systems students with basic programming familiarity. Participants will be randomly assigned to one of the three experimental conditions which are described below.

Task and Procedure. The experimental task will involve software development tasks using a hypothetical programming language created for the experiment. We will use a hypothetical language to minimize variability in prior language-specific expertise and to establish a more comparable baseline across participants. We select coding as the application domain because GenAI is widely used and particularly capable in software-related tasks [19].

Participants will complete two coding tasks. Task 1 (training task) will require implementing a function that computes Fibonacci numbers. Task 2 (evaluation task) will require implementing a function that computes Tribonacci numbers, which is structurally similar yet non-identical and comparable to support task completion. In the *no offloading* condition (control), participants receive a static text manual of the programming language to assist performance. In the *substitutive offloading* treatment, participants use a GenAI system to generate and complete the coding solution, fully offloading the task to the system. In the *constructive offloading* treatment, participants first develop an initial solution using only the manual and are subsequently allowed to use GenAI for feedback and improvements of their code.

Measures. We propose to measure cognitive offloading using alpha-band event-related spectral perturbation (ERSP) derived from time–frequency EEG analysis. Following prior research, we focus on task-related desynchronization in the upper alpha band (10–13 Hz), as reductions in alpha power reliably indicate increased cortical activation and memory-related processing in parietal and frontal regions [31, 32]. Consistent with established IS research, we will analyze alpha ERSP during task engagement to assess systematic differences in neural activation associated with offloading behavior [29, 33]. Higher alpha activity indicates lower cognitive load, thereby allowing researchers to objectively capture the extent to which users offload cognitive processes to GenAI across the no-offloading, constructive offloading, and substitutive offloading conditions.

We will measure task performance during Task 1 through the execution time and functional correctness of the code provided. Accordingly, we will measure skill development via a transfer-learning test using Task 2 performed without any external support. Skill development will be operationalized through execution time and functional correctness of the submitted solution. We will control for age, gender, handedness (left/right), education, AI literacy, algorithm aversion, prior software development skills, and perceived task difficulty.

Expected Contributions

This research contributes to the information systems literature in three ways. First, it advances an account of the underlying mechanisms linking GenAI use to skill development by conceptualizing cognitive offloading as the underlying mechanism. Second, it develops a conceptualization that distinguishes how cognitive offloading affects task performance and skill development across offloading modalities, clarifying when GenAI enhances skill development versus when it introduces dependencies. Third, this study contributes to NeuroIS by providing objective, real-time neurophysiological evidence of how cognitive offloading to GenAI affects users' cognitive processes during skill development, which cannot be captured through traditional self-report measures [34].

References

1. Eloundou T, Manning S, Mishkin P, Rock D (2024) GPTs are GPTs: Labor market impact potential of LLMs. Science 384:1306–1308. https://doi.org/10.1126/science.adj0998
2. Cui KZ, Demirer M, Jaffe S, et al (2026) The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. Management Science mnsc.2025.00535. https://doi.org/10.1287/mnsc.2025.00535
3. Brynjolfsson E, Li D, Raymond L (2025) Generative AI at Work. The Quarterly Journal of Economics 140:889–942. https://doi.org/10.1093/qje/qjae044
4. Chen Z, Chan J (2024) Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise. Management Science 70:9101–9117. https://doi.org/10.1287/mnsc.2023.03014
5. Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Science 381:187–192. https://doi.org/10.1126/science.adh2586
6. Diebel C, Goutier M, Martin Adam, Benlian A (2025) The Price of AI Assistance: The Undermining Effect of AI-Generated Code on Developers' Procedural Knowledge. In: CIS 2025 Proceedings. 2.
7. Risko EF, Medimorec S, Chisholm J, Kingstone A (2014) Rotating With Rotated Text: A Natural Behavior Approach to Investigating Cognitive Offloading. Cognitive Science 38:537–564. https://doi.org/10.1111/cogs.12087
8. Risko EF, Gilbert SJ (2016) Cognitive Offloading. Trends in Cognitive Sciences 20:676–688. https://doi.org/10.1016/j.tics.2016.07.002
9. Riedl R, Léger P-M (2016) Fundamentals of NeuroIS: Information Systems and the Brain, 1st ed. 2016. Springer, Berlin Heidelberg
10. Sweller J (1988) Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science 12:257–285. https://doi.org/10.1207/s15516709cog1202_4
11. 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
12. Das A (2003) Knowledge and Productivity in Technical Support Work. Management Science 49:416–431. https://doi.org/10.1287/mnsc.49.4.416.14419
13. Kim Y, Krishnan R, Argote L (2012) The Learning Curve of IT Knowledge Workers in a Computing Call Center. Information Systems Research 23:887–902. https://doi.org/10.1287/isre.1110.0401
14. Rinta-Kahila T, The University of Queensland Business School, Penttinen E, et al (2023) The Vicious Circles of Skill Erosion: A Case Study of Cognitive Automation. JAIS 24:1378–1412. https://doi.org/10.17705/1jais.00829
15. Yi MY, Davis FD (2003) Developing and Validating an Observational Learning Model of Computer Software Training and Skill Acquisition. Information Systems Research 14:146–169. https://doi.org/10.1287/isre.14.2.146.16016
16. Anderson JR (1982) Acquisition of cognitive skill. Psychological Review 89:369–406. https://doi.org/10.1037/0033-295X.89.4.369
17. Subramani M, Wagle M, Microsoft, et al (2021) Capability Development through Just-in-Time Access to Knowledge in Document Repositories: A Longitudinal Examination of Technical Problem Solving. MISQ 45:1287–1308. https://doi.org/10.25300/MISQ/2021/15635

18. Pan S, Tianjin University, Kwak D-H, et al (2024) Roles of Feedback and Phishing Characteristics in Antiphishing Training Performance: Perspectives of Goal Setting and Skill Acquisition. JAIS 25:1037–1078. https://doi.org/10.17705/1jais.00854
19. Bubeck S, Chandrasekaran V, Eldan R, et al (2023) Sparks of Artificial General Intelligence: Early experiments with GPT-4
20. Feuerriegel S, Hartmann J, Janiesch C, Zschech P (2024) Generative AI. Bus Inf Syst Eng 66:111–126. https://doi.org/10.1007/s12599-023-00834-7
21. Alavi M, Leidner DE, Mousavi R (2024) A Knowledge Management Perspective of Generative Artificial Intelligence. JAIS 25:1–12. https://doi.org/10.17705/1jais.00859
22. Chiu G, Gilbert SJ (2024) Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions. Quarterly Journal of Experimental Psychology 77:1295–1311. https://doi.org/10.1177/17470218231199977
23. Gilbert SJ (2015) Strategic offloading of delayed intentions into the external environment. Quarterly Journal of Experimental Psychology 68:971–992. https://doi.org/10.1080/17470218.2014.972963
24. Boldt A, Gilbert SJ (2019) Confidence guides spontaneous cognitive offloading. Cogn Research 4:45. https://doi.org/10.1186/s41235-019-0195-y
25. Richmond LL, Taylor RG (2025) The benefits and potential costs of cognitive offloading for retrospective information. Nat Rev Psychol 4:312–321. https://doi.org/10.1038/s44159-025-00432-2
26. Grinschgl S, Papenmeier F, Meyerhoff HS (2021) Consequences of cognitive offloading: Boosting performance but diminishing memory. Quarterly Journal of Experimental Psychology 74:1477–1496. https://doi.org/10.1177/17470218211008060
27. Sparrow B, Liu J, Wegner DM (2011) Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. Science 333:776–778. https://doi.org/10.1126/science.1207745
28. Storm BC, Stone SM (2015) Saving-Enhanced Memory: The Benefits of Saving on the Learning and Remembering of New Information. Psychol Sci 26:182–188. https://doi.org/10.1177/0956797614559285
29. Moravec PL, Minas RK, Dennis AR (2019) Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All1. MIS Quarterly 43:1343–1360. https://doi.org/10.25300/MISQ/2019/15505
30. Lakhiwal A, Bala H, Léger P-M (2023) Ambivalence Is Better than Indifference: A Behavioral and Neurophysiological Assessment of Ambivalence in Online Environments. MIS Quarterly 47:705–732. https://doi.org/10.25300/MISQ/2022/17123
31. Klimesch W (2012) Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences 16:606–617. https://doi.org/10.1016/j.tics.2012.10.007
32. Makeig S (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalography and Clinical Neurophysiology 86:283–293. https://doi.org/10.1016/0013-4694(93)90110-H
33. Minas RK, Potter RF, Dennis AR, et al (2014) Putting on the Thinking Cap: Using NeuroIS to Understand Information Processing Biases in Virtual Teams. Journal of Management Information Systems 30:49–82. https://doi.org/10.2753/MIS0742-1222300403
34. Dimoka A, Banker RD, Benbasat I, et al (2012) On The Use of Neurophysiological Tools in Is Research: Developing A Research Agenda for Neurois1. MIS Quarterly 36:679–702. https://doi.org/10.2307/41703475

Neurodiversity and Technostress: Towards a Multimodal Research Design for Evaluating Subjective, Physiological, and Behavioral Responses

Lisa van den Heuvel[1](✉), Igor Ivkić[1,2], René Riedl[3,4]

[1]*University of Applied Science Burgenland, Eisenstadt, AT*

[2]*Lancaster University, Lancaster, UK*

[3]*University of Applied Sciences Upper Austria, Steyr, AT*

[4]*Johannes Kepler University, Linz, AT*

2410781017@hochschule-burgenland.at, i.ivkic@lancaster.ac.uk, rene.riedl@fh-steyr.at

Abstract. Digitalization has transformed modern work by increasing efficiency while also introducing new forms of strain. Technostress (TS) describes subjective, physiological, and behavioral stress responses related to digital technology use. Existing TS research has predominantly focused on neurotypical populations and rarely integrates multiple stress dimensions within a single design. This paper addresses these gaps by proposing a controlled experimental research design that systematically compares neurodivergent and neurotypical individuals under standardized digital stress conditions. The proposed design combines structured and unstructured digital tasks with a multimodal measurement approach covering subjective perceptions, physiological activation, and observable interaction behavior. By integrating neurodiversity into TS research, the paper contributes to a more differentiated understanding of digital stress and provides a methodological approach for more inclusive digital work design.

Keywords: Technostress · Neurodiversity · Digital Stress · Research Design · Stress Physiology · Cognitive Workload · Human–Computer Interaction

Introduction

Digital technologies have become a central part of modern work environments. Information and communication technologies enable new forms of collaboration, increase flexibility, and structure core work processes [1, 2, 3]. At the same time, the growing reliance on digital systems is associated with increasing experiences of overload, interruption, and expectations of constant availability. These phenomena are commonly summarized under the term Technostress (TS), also referred to as digital stress [1, 2, 4, 5].

TS describes stress reactions that arise in response to the use of digital technologies, particularly when these technologies are experienced as complex, interruptive, or difficult to control [1, 2]. Prior research has shown that TS can negatively affect well-being, job satisfaction, and performance, and may also influence work behavior and organizational outcomes [2, 4-6]. At the same time, recent studies emphasize that digital technologies can also act as supportive resources, for example by enabling autonomy or improving work processes, highlighting the dual nature of technology as both a stressor and a resource [6-8].

This dual role highlights that TS should not be understood as a purely negative phenomenon, but rather as the result of an interaction between technological demands, available resources, and individual characteristics [8]. In this context, digital stress depends not only on technological features but also on how individuals perceive and process these demands. However, existing research has primarily focused on neurotypical (NT) populations, largely overlooking differences in cognitive and sensory processing across individuals.

Open questions remain regarding how different groups of individuals perceive and process digital stressors. In particular, neurodivergent (ND) individuals, such as those with attention-deficit/hyperactivity disorder (ADHD), autism spectrum conditions, or dyslexia, often show distinct patterns in attention regulation, executive functioning, and sensory processing [10-13]. As these characteristics are highly relevant in digital work environments, ND individuals may experience and process digital stressors differently compared to NT individuals.

Despite the increasing relevance of neurodiversity in the workplace, empirical research that systematically examines these differences in the context of TS remains limited [1, 2, 4]. Furthermore, existing TS research often focuses on either subjective, physiological, or behavioral indicators, while multimodal approaches have only rarely been applied to compare different user groups within a single research design [15-19]. This constitutes an important limitation, as stress is a multi-layered phenomenon in which subjective perception, physiological activation, and behavioral responses may diverge [15].

To overcome this limitation, this paper proposes a controlled experimental research design that systematically compares ND and NT individuals under standardized digital stress conditions. By integrating subjective, physiological, and behavioral measures within a multimodal framework, the paper aims to contribute to a more differentiated understanding of TS. This design enables the analysis of how ND and NT individuals respond to digital stressors and provides a basis for designing more inclusive digital work environments.

Background and Research Gap

This section summarizes key concepts related to TS and neurodivergence, identifies the research gap addressed in this paper, reviews TS in digital workplaces, examines how neurodivergence influences information processing and stress responses, and highlights the need for a multimodal research approach.

Technostress in Digital Work Contexts

Research on TS primarily originates from the fields of information systems (IS) and occupational psychology. Previous studies have identified core categories of technology-related stressors, including techno-overload, techno-invasion, techno-complexity, and techno-uncertainty [1, 2, 4]. These stressors describe situations in which individuals experience excessive demands, blurred work–life boundaries, difficulties in using technologies, or constant technological change.

More recent research suggests that traditional TS frameworks may not fully capture the evolving nature of digital work environments. Emerging stressors such as system unreliability or increased monitoring, or constant tool switching have been identified as additional sources of strain [8, 20]. In response, newer measurement approaches, such as the digital stressors scale (DSS), conceptualize TS as a multidimensional construct that reflects a broader range of technology-related demands [21].

TS has been associated with a variety of negative outcomes, including cognitive overload, reduced job satisfaction, decreased performance, and health-related impairments [1, 2, 4, 6]. It has also been linked to negative organizational outcomes, such as changes in work behavior and increased tendencies toward knowledge hiding [22]. Experimental research further indicates that digital stressors are associated with physiological stress responses, including changes in heart rate variability (HRV), blood pressure, and cortisol levels [15, 17, 18, 23, 24]. These physiological responses reflect activation of the autonomic nervous system and provide objective indicators of stress.

At the same time, digital technologies should not be viewed exclusively as sources of stress but can also function as supportive resources. Depending on their design, the context of use, and individual characteristics, digital technologies can be experienced as a supportive resource, for example, when they enhance autonomy or facilitate work processes [6, 7]. Concepts such as techno-eustress describe situations in which technology use is experienced as motivating or performance-enhancing [6], while techno-relief refers to the potential of digital tools to reduce stress and support work processes [25]. In this context, digital technologies can act both as hindrance stressors (e.g., interruptions, overload) and as challenge stressors (e.g., stimulating tasks, autonomy) [26].

Given the multidimensional and context-dependent nature of TS, researchers have emphasized the importance of integrating subjective, physiological, and behavioral indicators, as these may capture different but complementary aspects of stress [16, 27]. However, and to the best of our knowledge, such multimodal approaches have rarely been applied in a comparative context across different user groups.

Neurodivergence and Stress Processing

Alongside TS research, the concept of neurodiversity has gained increasing attention in organizational and information systems contexts. Neurodiversity refers to natural variations in how individuals perceive, process, and respond to information, including conditions such as ADHD, autism spectrum conditions, and dyslexia [10–13]. Recent IS research also argues that neurodiversity should be considered more systematically because neurodiversity facets can influence cognition, emotion, decision-making, and user behavior, and because ignoring such differences may reduce the inclusivity and explanatory power of IS research [28].

While neurodiversity encompasses a broad spectrum of conditions, this study focuses on selected forms of neurodivergence that are particularly relevant in digital work environments. Specifically, ADHD, autism spectrum conditions, and dyslexia are considered, as they are associated with differences in attention regulation, executive functioning, and sensory processing that are highly relevant in technology-mediated work settings [29].

Prior research suggests that individuals with ADHD may experience difficulties in sustained attention, impulse control, and executive functioning, which can affect their ability to manage multiple tasks, maintain focus, and resist distractions in digital environments [14]. Autism spectrum conditions are associated with differences in sensory processing and information filtering, which may increase sensitivity to environmental stimuli and contribute to sensory overload in complex or dynamic digital interfaces [14, 30]. Dyslexia, characterized by persistent difficulties in reading and processing written language, may influence how individuals interact with text-intensive digital systems and increase cognitive effort in information processing tasks [14]

These characteristics are highly relevant in digital work contexts and may lead to differences in subjective, physiological, and behavioral stress responses. For example, differences in attention regulation may be reflected in interaction patterns such as task-switching frequency, response times, or error rates, while sensory sensitivity may influence visual attention and gaze behavior in complex interfaces [18].

At the same time, workplace research highlights that ND individuals can contribute specific strengths, such as attention to detail, pattern recognition, or problem-solving abilities, particularly in supportive environments [10-13]. However, despite these insights, quantitative and experimental research that directly compares ND and NT individuals in controlled digital work scenarios remains limited.

Research Gap and Research Questions

While prior research has provided valuable insights into TS and, separately, into neurodiversity in the workplace, these research streams have largely evolved independently. TS research has predominantly focused on NT populations, whereas neurodiversity research has mainly examined cognitive characteristics, inclusion, and workplace strengths and challenges. As a result, there is limited empirical evidence on how ND and NT individuals differ in their perception of and response to digital stressors in work-related contexts.

In addition, TS is increasingly conceptualized as a multidimensional and context-dependent phenomenon. However, many studies focus on isolated aspects, such as self-reported perceptions or individual physiological indicators. Although multimodal approaches have been proposed in prior research, they have rarely been applied to systematically compare ND and NT individuals in controlled digital work environments [27]. This is particularly relevant, as subjective perception, physiological activation, and observable behavior may diverge.

Furthermore, existing research is predominantly based on survey-based or cross-sectional designs, while controlled experimental studies in realistic digital work scenarios remain limited. This constrains the ability to examine causal relationships and to compare how different user groups respond to controlled digital stress conditions.

Against this background, this study proposes a controlled experimental research design that systematically compares ND and NT individuals under controlled digital stress conditions. The design integrates subjective perceptions, physiological stress responses, and observable interaction behavior in a multimodal approach.

Based on this approach, the following research questions (RQ) are proposed:

- **RQ1:** To what extent do ND and NT individuals differ in their subjective perception of digital stressors in a controlled digital work scenario?
- **RQ2:** To what extent do ND and NT individuals differ in their physiological stress reactions (heart rate, heart rate variability) under digital stress conditions?
- **RQ3:** To what extent do ND and NT individuals differ in observable work and interaction behavior under digital stress conditions?

Methods

To answer the proposed research questions, this section presents the methodological framework of the study. It describes the experimental design, participant selection, procedure, measurement approach, and planned data analysis. The aim is to provide a coherent and transparent description of how differences in TS responses between ND and NT individuals can be examined in a controlled setting.

Research Design

This paper proposes a controlled experimental research design to systematically investigate how ND and NT individuals perceive and respond to digital stressors. To capture group-specific and context-dependent effects, the design combines a between-subjects factor and a within-subjects factor. Neurodivergence (ND and NT groups) serves as the between-subjects factor, while participants complete different digital task types as within-subject conditions.

This approach allows comparison of stress responses between groups, while also examining how the same individuals react across different digital work contexts. In addition, the use of two task types supports a more differentiated interpretation of results and improves the potential generalizability of the proposed design.

An overview of the proposed experimental flow is provided in Figure 1. The procedure consists of a baseline phase, a task phase with counterbalanced task order, an immediate post-task phase, and a follow-up assessment capturing perceptions of digital stress in everyday work settings.

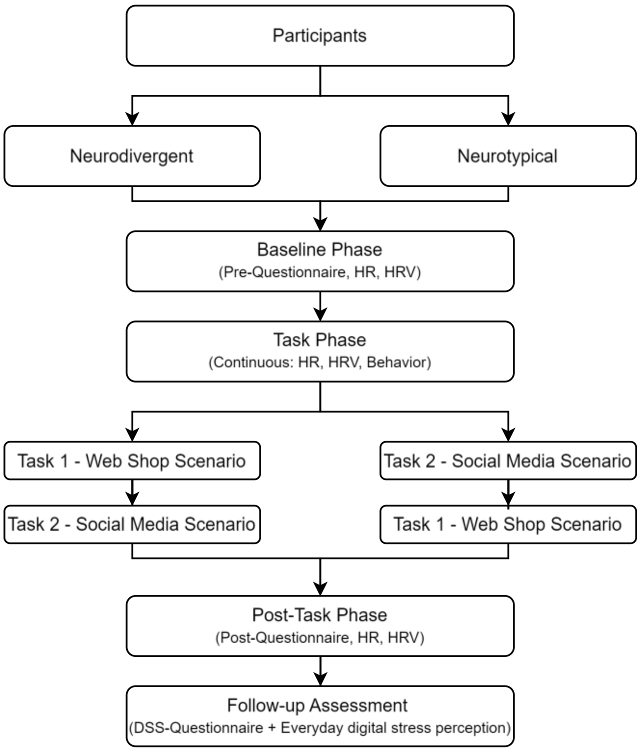


Fig. 10. Overview of the proposed experimental design

Participants

As shown in Fig. 1, the proposed study aims to include both ND and NT participants. ND status would be operationalized through self-report and would focus on forms of neurodivergence that are particularly relevant to digital work environments, including ADHD, autism spectrum conditions, and dyslexia. Because the study is concerned with work-related digital stress, the target sample should consist of adults who regularly use digital tools in work-related or closely comparable settings.

Participants could be recruited through online channels, professional networks, and, where appropriate, university populations with relevant digital work experience. In addition to group membership, demographic and contextual information, such as age, gender, education and prior experience with digital tools, would be collected to account for potential confounding influences and ensure comparability between groups [31].

With regard to sample size, the proposed design should aim for at least 50 participants per group as a lower bound for meaningful group comparisons. Larger samples would be preferable to increase statistical power but may not be feasible within practical constraints. Because this paper presents a research design rather than an implemented study, this number should be understood as a design-oriented benchmark rather than as a finalized recruitment target.

Experimental Tasks

To approximate realistic digital work environments, the proposed design includes two experimental tasks reflecting different forms of digital interaction and stress exposure:

Task 1: Structured task – Web shop scenario. The first task is a structured, goal-oriented scenario based on a web shop environment. Participants are asked to complete a clearly defined task, such as placing an online order. During task execution, controlled digital stressors are introduced, including interruptions, changing requirements, system delays, temporary malfunctions or crashes, and time pressure. This task is intended to represent process-driven digital work with explicit goals and external constraints.

Task 2: Unstructured task – Social media scenario. The second task is a more unstructured and dynamic scenario based on a social-media-like environment. Participants perform loosely defined activities, such as interacting with content, writing short responses, maintaining a chat conversation, and switching between parallel activities. In contrast to the web shop task, this scenario is intended to reflect less structured digital environments characterized by continuous information flow, multitasking, and higher demands on cognitive flexibility.

Across both tasks, TS is induced through experimentally controlled stressors embedded in the digital environment, including interruptions, multitasking demands, time pressure, and system-related issues. This approach allows digital stress to be represented in a controlled yet ecologically plausible manner.

To reduce potential order effects, the sequence of tasks would be systematically varied across participants. One subgroup would start with the structured task and then complete the unstructured task, while the other subgroup would follow the reverse order. This counterbalancing procedure is intended to minimize order effects such as fatigue, learning, task familiarity, and potential carry-over effects between tasks.

Procedure

The following phases structure the experimental procedure and capture stress responses across baseline, task execution, post-task assessment, and follow-up self-report. To reduce expectation effects, participants are not explicitly informed that stress represents the main focus of the study. Instead, a neutral description of the experiment is provided to minimize anticipatory stress and potential response biases.

Baseline Measurements Phase. The first phase would consist of a short baseline period under resting conditions. During this phase, heart rate (HR) and HRV would be recorded to establish individual physiological reference values. In addition, participants would complete a brief pre-task questionnaire assessing their current condition, including perceived energy level, sleep quality, recent physical activity, caffeine or nicotine intake, and medication use. This information is intended to document factors that may influence physiological stress responses independently of the experimental tasks.

Task Phase. In the second phase, participants would complete both experimental tasks in counterbalanced order. During task execution, physiological and behavioral data would be recorded continuously. This setup would allow stress responses to be examined in relation to specific task events, such as interruptions, delays, or system errors.

Post-Task Phase. The third phase would consist of an immediate post-task assessment. Physiological recording would continue for a short period after task completion in order to capture short-term recovery processes. Participants would also complete a post-task questionnaire assessing perceived stress, task difficulty, and subjective experience during the tasks. Following task completion, participants would be debriefed about the purpose of the experiment and given the opportunity to ask questions.

Follow-up Assessment. In addition, the design proposes a follow-up self-report assessment after a defined interval. This follow-up would include standardized TS-related instruments, such as the DSS [21], as well as semi-structured questions on how participants perceive digital stress in their everyday work contexts. The purpose of this follow-up is to complement the immediate laboratory-based assessment with a less situation-specific perspective on digital stress in daily work life.

Measures

To reflect the multidimensional nature of TS, the proposed design follows a multimodal measurement approach that integrates the following subjective, physiological, and behavioral indicators:

Subjective stress. Subjective stress would be assessed at several points in time. Pre-task measures would capture the participants' initial condition before stress induction. Immediate post-task measures would assess perceived stress, task difficulty, and subjective experience during the experimental tasks. In addition, the follow-up assessment would be used to capture more stable and context-related perceptions of digital stress in everyday work. For this assessment, standardized instruments such as the DSS [21] could be combined with Likert-scale items and selected semi-structured questions to capture subjective experiences more comprehensively.

Physiological measures. Physiological measures would include HR and HRV, recorded using a wearable chest-strap device. Measurements would be taken during baseline, throughout task execution, and during the immediate post-task phase. Baseline values are important because they provide an individual physiological reference point and allow stress-related changes to be interpreted relative to each participant's resting state. Physiological responses reflect activation of the autonomic nervous system and provide objective indicators of stress. In particular, HRV-based metrics such as the root mean square of successive differences (RMSSD) could be analyzed to capture short-term variations in autonomic regulation, especially parasympathetic activity [32].

Behavioral data. Behavioral data are collected during task execution using interaction-tracking methods, including mouse movements, click behavior, response times, and task-switching patterns. These measures help capture observable responses to digital stressors, such as increased hesitation, error-proneness, or inefficient interaction patterns. If feasible, eye-tracking can be integrated to assess visual attention and cognitive processing during task performance. Metrics such as fixation duration and gaze transitions provide additional insights into attentional strategies and cognitive load in complex digital environments.

Data Analysis

The proposed data analysis follows the multimodal research design and integrates subjective, physiological, and behavioral indicators. Group comparisons between ND and NT can be conducted using statistical methods appropriate for the mixed design, such as repeated measures analysis of variance (ANOVA), to assess within-subject effects across task types, between-group differences, and interaction effects between group membership and task condition.

Physiological data can be analyzed relative to baseline values to assess changes in activation during the task phase and short-term recovery in the post-task phase. Depending on data structure and quality, mixed-effects models may be considered, particularly in the presence of repeated or unbalanced observations.

In addition, correlation and regression analyses can examine relationships between subjective stress perception, physiological responses, and behavioral indicators, allowing the evaluation of convergence or divergence between perceived and objectively measured stress responses. Such discrepancies are especially relevant in TS, where physiological activation, subjective appraisal, and behavior may not always align.

Finally, exploratory analyses can consider whether contextual or individual variables, such as prior experience with digital tools, are associated with stronger or weaker stress responses. These analyses are exploratory rather than confirmatory and can serve as a basis for future research.

Ethical Considerations

The proposed study induces and measures TS in a controlled laboratory setting and therefore requires clear ethical safeguards to protect participants' well-being and autonomy. Participants would receive information on the procedure and measurement methods, provide written informed consent, and be free to withdraw at any time without disadvantage.

Although the tasks are designed to induce digital stress, the procedure would remain within the range of typical workplace-related strain rather than excessive burden. Participants would be monitored throughout the experiment, and the procedure would be stopped if signs of significant distress occur. After the study, participants would be debriefed and given the opportunity to discuss their experience. All collected data would be stored pseudonymously and used solely for research purposes in compliance with data protection regulations.

Discussion and Outlook

This paper addresses the underexplored intersection of TS and neurodiversity by proposing a controlled experimental research design that combines subjective, physiological, and behavioral measures. The proposed design responds to three limitations in prior research: the limited integration of neurodiversity into TS research, the limited use of multimodal approaches in comparative studies across user groups, and the scarcity of controlled experimental studies in realistic digital work contexts.

By systematically comparing ND and NT individuals, the proposed design acknowledges that digital stress should not be treated as a homogeneous phenomenon. Combining self-reports with physiological and behavioral indicators supports a more differentiated understanding of stress responses, especially where perceived stress and objective indicators diverge.

Several limitations must be acknowledged. First, the proposed design is based on a laboratory setting, which may reduce ecological validity compared to real-world work environments [33]. Second, ND groups are inherently heterogeneous, and the selected conditions cannot represent the full spectrum of neurodiversity. Third, the proposed sample size and recruitment feasibility may constrain practical implementation.

Future research could extend this approach through field studies, larger and more differentiated samples, and the inclusion of contextual factors such as coping strategies, digital work experience, or workplace design.

Conclusion

Digitalization has fundamentally transformed modern work environments, while also introducing new forms of strain commonly described as TS. Within this research field, individual differences in how digital stressors are perceived and processed remain insufficiently understood, particularly with regard to neurodiversity. Existing TS research has largely focused on neurotypical populations and has rarely combined subjective, physiological, and behavioral perspectives within a single design.

To address these limitations, this paper proposes a controlled experimental research design that systematically compares ND and NT individuals under standardized digital stress conditions. By integrating structured and unstructured task scenarios with a multimodal measurement approach, the design enables a differentiated analysis of stress responses across multiple dimensions.

The key contribution of this work lies in explicitly linking neurodiversity to TS research and demonstrating the importance of considering heterogeneous cognitive and sensory processing patterns when studying digital stress. The proposed approach highlights that TS should not be treated as a uniform phenomenon, but rather as a complex interaction between technological demands and individual characteristics.

Future research can build on this design to conduct empirical studies, extend the approach to real-world work settings, and further explore how digital environments can be designed to support diverse user needs and promote more inclusive and cognitively adaptive workplaces.

References

1. Ragu-Nathan, T.S., Tarafdar, M., Ragu-Nathan, B.S., Tu, Q.: The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. *Information Systems Research* 19(4), pp. 417–433 (2008)
2. Tarafdar, M., Tu, Q., Ragu-Nathan, B.S., Ragu-Nathan, T.S.: The Impact of Technostress on Role Stress and Productivity. *Journal of Management Information Systems* 24(1), pp. 301–328 (2007)
3. Salazar-Concha, C., Ficapal-Cusi, P., Boada-Grau, J., Camacho, L.J.: Analyzing the Evolution of Technostress: A Science Mapping Approach. *Heliyon* 7(4), e06726 (2021)
4. Tarafdar, M., Tu, Q., Ragu-Nathan, T.S., Ragu-Nathan, B.S.: Crossing to the Dark Side: Examining Creators, Outcomes, and Inhibitors of Technostress. *Communications of the ACM* 54(9), pp. 113–120 (2011)
5. Ayyagari, R., Grover, V., Purvis, R.: Technostress: Technological Antecedents and Implications. MIS Quarterly 35(4), 831–858 (2011)
6. Tarafdar, M., Cooper, C.L., Stich, J.-F.: The Technostress Trifecta: Techno-Eustress, Techno-Distress and Design. *Information Systems Journal* 29(1), pp. 6–42 (2019)
7. Bakker, A.B., Demerouti, E.: Job Demands–Resources Theory: Taking Stock and Looking Forward. *Journal of Occupational Health Psychology* 22(3), pp. 273–285 (2017)
8. Fischer, T., Riedl, R.: On the stress potential of an organisational climate of innovation: a survey study in Germany. Behaviour & Information Technology, 41(4), pp. 805–826 (2022).
9. Fischer, T., Riedl, R.: Theorizing Technostress in Organizations: A Cybernetic Approach. In: Thomas, O., Teuteberg, F. (eds.) Proceedings of the 12th International Conference on Wirtschaftsinformatik (WI 2015), Osnabrück, pp. 1453–1467 (2015)
10. Austin, R.D., Pisano, G.P.: Neurodiversity as a Competitive Advantage. *Harvard Business Review* (2017)
11. Lauder, K.: Towards Neuro-Inclusive Workplaces: Insights from Neurodivergent Individuals in a Conservation-Based Organization. *Journal of Organizational Psychology* 24(4) (2024)
12. Weber, C., Krieger, B., Häne, E., Yarker, J., McDowall, A.: Physical Workplace Adjustments to Support Neurodivergent Workers: A Systematic Review. *Applied Psychology* 73, pp. 910–962 (2022)
13. Rollnik-Sadowska, E., Grabinska, V.: Managing Neurodiversity in Workplaces: A Review and Future Research Agenda for Sustainable Human Resource Management. *Sustainability* 16, 6594 (2024)
14. Helmold, M., Martensen, M., Hummel, F.: Neurodiversity in organisations: Effective strategies for leadership and management. World Scientific (2025)
15. Riedl, R.: On the biology of technostress: Literature review and research agenda. ACM SIGMIS Database for Advances in Information Systems 44(1), 18–55 (2013)
16. Fischer, T., Riedl, R.: Technostress Research: A Nurturing Ground for Measurement Pluralism? Communications of the Association for Information Systems 40(17), pp. 375-401 (2017)
17. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress aus einer neurobiologischen Perspektive: Systemabsturz führt bei Computerbenutzern zu einem Anstieg des Stresshormons Kortisol. *Wirtschaftsinformatik* 54(2), pp. 59–68 (2012)
18. Kim, S.Y., Park, H., Kim, H., Kim, J., Seo, K.: Technostress Causes Cognitive Overload in High-Stress People: Eye Tracking Analysis in a Virtual Kiosk Test. *Information Processing & Management* 59(6), 103093 (2022)

19. Tams, S., Thatcher, J.B., Grover, V.: Concentration, competence, confidence, and capture: An experimental study of age, interruption-based technostress, and task performance. Journal of the Association for Information Systems 19(9) (2018)
20. Fischer, T., Pehböck, A., Riedl, R.: Is the Technostress Creators Inventory Still an Up-To-Date Measurement Instrument? Results of a Large-Scale Interview Study. In Tagungsband der 14. Internationalen Tagung Wirtschaftsinformatik, pp. 1834-1845 (2019)
21. Riedl, R., Fischer, T., Reuter, M. Measuring digital stress in the workplace context. Informatik Spektrum 46, pp. 235–239 (2023)
22. Shen, B., Kuang, Y.: Assessing the Relationship Between Technostress and Knowledge Hiding: A Moderated Mediation Model. *Data and Information Management* 6, 100002 (2022)
23. Baumgartner, D., Fischer, T., Riedl, R., Dreiseitl, S.: Analysis of Heart Rate Variability (HRV) Feature Robustness for Measuring Technostress. In: *Lecture Notes in Information Systems and Organisation*, vol. 29, pp. 221–228. Springer (2019)
24. Fischer, T., Halmerbauer, G., Meyr, E., Riedl, R.: Blood Pressure Measurement: A Classic of Stress Measurement and Its Role in Technostress Research. In: Davis, F., Riedl, R., vom Brocke, J., Léger, P.M., Randolph, A. (eds.) *Information Systems and Neuroscience*, LNISO, vol. 25. Springer, pp. 25-35 (2017)
25. Saigot, M.: Unveiling Technorelief: Enhancing Neurodiverse Collaboration With Digital Capabilities. In: Bui, T.X. (ed.) Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS 2024), pp. 6921–6931 (2024).
26. LePine, M. A.: The Challenge-Hindrance Stressor Framework: An Integrative Conceptual Review and Path Forward. Group & Organization Management, 47(2), 223-254. (2022)
27. Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., Grover, V.; 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(10), 1 (2014)
28. Turel, O., Bechara, A.: Time to talk about it: Neurodiversity of information systems users in leisure settings. Journal of the Association for Information Systems 25(5), 1343–1371 (2024)
29. Soubiran, Vera. Neurodiversität in Unternehmen : Herausforderungen und Chancen Für Neurodiverse Teams, Springer Berlin / Heidelberg (2025)
30. Robertson, C., Baron-Cohen, S.: Sensory perception in autism. Nature reviews Neuroscience. 18 (2017)
31. Javor A, Ransmayr G, Struhal W., Riedl R.: Parkinson Patients' Initial Trust in Avatars: Theory and Evidence. PLoS ONE 11(11), e0165998 (2016)
32. Shaffer F., Ginsberg JP.: An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health 5, 258 (2017)
33. Balapour, A., Riedl, R.: Ecological Validity in NeuroIS Research: Theory, Evidence, and a Roadmap for Future Studies, Journal of the Association for Information Systems, 26(1), 9-65 (2025)

Brain–Body Responses to Workplace Discrimination in the Context of Inclusive Decision-Making

Sareli Gibaja[1], Ravi Kanth Kosuru[2], Leidy Cubillos, Ksenia Keplinger[1], Mathias Vukelić[3]

[1] Max Planck Institute for Intelligent Systems, Stuttgart, Germany
{sgibaja, kkeplinger}@is.mpg.de
[2] University of Stuttgart, Institute of Human Factors and Technology Management IAT, Applied Neurocognitive Systems, Stuttgart, Germany
ravi-kanth.kosuru@iat.uni-stuttgart.de
[3] Applied Neurocognitive Systems, Fraunhofer Institute for Industrial Engineering IAO, Stuttgart, Germany
mathias.vukelic@iao.fraunhofer.de

Abstract. Why do some decision-makers respond to workplace discrimination with inclusive action, while others remain unaffected? Addressing this question requires moving beyond self-report to examine affective and physiological processes underlying leadership decision-making. Grounded in Affective Events Theory, this study investigates whether discriminatory workplace scenarios trigger affective and neural responses that influence decisions. Sixty participants viewed neutral or discriminatory workplace videos while multimodal neurophysiological signals, including electroencephalography, electrocardiography, and electrooculography, were recorded. They then performed hiring tasks assessing inclusive choices. Self-report analyses indicate differences between conditions, accompanied by autonomic effects. Ongoing analyses assess whether discriminatory scenarios elicit stronger affective-interoceptive engagement, and whether these responses predict inclusive decision-making. By integrating brain-body measures with behavioral outcomes, we aim to provide a process-level view of inclusive decision-making and identify neurophysiological markers associated with it. These findings contribute to research on decision processes and inform future work on adaptive decision environments and bias-aware support systems.

Keywords: Inclusive decision-making · Affective Events Theory · Multimodal neurophysiological assessment · EEG · ECG · EOG

Introduction

Decision-making in organizational contexts often involves evaluating socially and morally salient information [1]. Inclusive decision-making is shaped not only by candidate qualifications but also by how decision-makers process emotional cues that influence judgment and choice [2]. Such emotionally salient events may alter attentional and evaluative processes, influencing how fairness considerations are integrated into later evaluations [3].

Despite these advances, most research on inclusive leadership and biased decision-making relies on self-report measures [4, 5]. While informative, these approaches provide limited access to the real-time cognitive and affective processes through which workplace events shape decisions. Moreover, prior neurophysiological research has primarily focused on economic exchanges, trust, or abstract risk paradigms [6], leaving open how affective responses to discrimination-specific workplace events influence subsequent organizational decisions. This limitation is critical because workplace discrimination represents a socially and morally charged context in which affective responses are likely to influence fairness-related judgments and hiring decisions [7]. Accordingly, the present study addresses the following research question: To what extent do discriminatory workplace scenarios elicit affective and neurophysiological responses, and how are these responses related to subsequent inclusive hiring decisions?

To address this question, the present study adopts a multimodal neurophysiological approach. Participants are exposed to video-based workplace scenarios depicting discriminatory and neutral interactions while physiological responses are continuously recorded, including heart rate, heartbeat-evoked potentials, and blink-related EEG activity. These measures are combined with subjective affect ratings and a subsequent hiring decision task.

This study makes two contributions. First, it links real-time affective and physiological responses to inclusive decision-making in organizational contexts. Second, it extends prior NeuroIS research by examining discrimination-specific stimuli, moving beyond abstract economic paradigms to more ecologically valid workplace settings.

Related Work

This study builds on work within the NeuroIS paradigm, which integrates neuroscience tools to capture processes that individuals are unable or unwilling to report accurately, including deep emotional and moral evaluations [6, 8]. Applied to affective and decision processes, this approach enables continuous, real-time measurement of responses as they unfold during task execution [6].

Empirical work within the NeuroIS framework shows that neurophysiological signals capture underlying processes that influence decision-making in IS contexts. For example, Dimoka identified the neural correlates of trust in e-commerce decisions, linking activation in the caudate nucleus, anterior paracingulate cortex, and orbitofrontal cortex to reward expectations, prediction of others' behavior, and uncertainty in transactions [6, 9]. Building on this foundation, more recent research demonstrates that neurophysiological measures such as EEG and heart rate variability provide continuous, real-time indices of cognitive and affective states during decision-making. EEG studies show that variations in alpha and beta activity track working memory load and attentional engagement in complex decisions [10], while higher vagally mediated HRV is associated with better decision performance, particularly under risk and uncertainty [11]. Together, these findings establish that both neural and cardiac signals capture affective responses that exert measurable influence on decision behavior.

At the level of specific physiological markers, prior work shows that heartbeat-evoked potentials index affective-interoceptive processing [12], while heart rate reflects autonomic activation [13]. Attentional engagement during emotionally relevant events can be observed through ocular and electrophysiological indicators. Spontaneous blink rate and blink-related EEG activity are sensitive to sustained attentional engagement and cognitive workload under affective demands [14, 15]. Empirical work further shows that HEP dynamics vary with emotionally salient and arousal-related states and that emotional stimuli are associated with reduced blink rate. Affective arousal intensifies evaluations and shapes subsequent judgments [16].

Taken together, this study establishes that neurophysiological signals provide complementary indicators of affective and attentional processes that influence decision behavior. However, existing studies have largely focused on economic exchanges, technology use, or abstract risk paradigms. Existing NeuroIS work has only examined to a limited extent how neurophysiological responses to discrimination-specific stimuli translate into inclusive or non-inclusive hiring decisions in organizational contexts.

Theoretical Background and Hypotheses Development

Drawing on Affective Events Theory (AET) [7], discriminatory workplace interactions can be understood as affective events that trigger emotional appraisal, attentional shifts, and regulatory engagement during exposure. AET provides the overarching framework for the present hypotheses, explaining how workplace events elicit affective responses that shape subsequent evaluations and decisions. When perceived as violations of fairness and inclusion norms, such interactions may increase emotional salience and influence how fairness considerations are incorporated into later decisions [7]. These responses are reflected in physiological markers of autonomic arousal and attentional engagement, including heart rate, heartbeat-evoked potentials, and spontaneous blink rate.

H1: Exposure to discriminatory (vs. neutral) workplace scenarios will elicit greater neurophysiological arousal, indexed by elevated heart rate, modulation of heartbeat-evoked potential, and reduced spontaneous blink rate.

AET further posits that affective responses to workplace events influence subsequent judgments and behaviors [7]. Physiological arousal during affective events reflects the intensity of emotional appraisal and shapes how information is weighed in later evaluations. Specifically, stronger autonomic and interoceptive responses indicate heightened event salience, while sustained attentional engagement reflects deeper processing of fairness-relevant information. Affective responses are central to fair decision-making, with negative emotional reactions and associated arousal increasing the likelihood of norm enforcement behaviors such as rejection or punishment of unfairness [17, 18]. At the same time, affective engagement more broadly guides social decisions by increasing the salience and weighting of fairness-relevant information, thereby promoting prosocial and compensatory choices [19]. In the context of discriminatory scenarios, stronger affective responses are therefore expected to increase the salience of fairness considerations and promote corrective, inclusive decision-making.

H2: Greater neurophysiological arousal during exposure to discriminatory workplace scenarios will be positively associated with inclusive hiring decisions in the subsequent task.

While physiological markers capture the intensity of the affective response, subjective experience captures its evaluative quality, specifically how the event is appraised in terms of valence and perceived arousal. This distinction is important because valence reflects whether the event is perceived as a fairness violation and thus generates directional motivational pressure toward correction rather than mere activation. Within AET, the combination of negative appraisal and affective intensity determines both the direction and strength of the downstream behavioral effect. Accordingly, more negative valence and higher arousal are expected to translate into more inclusive hiring choices.

H3: More negative subjective valence and higher arousal in response to discriminatory workplace scenarios will be positively associated with more inclusive hiring decisions.

Method

Design and Participants

The study employed a within-subjects experimental design with multimodal physiological recording (EEG, ECG, EOG) to assess responses to workplace scenarios. This approach captures affective and attentional processes as they unfold in real time. By simultaneously recording cardiac, ocular, and cortical signals, the present design enables the assessment of distinct but complementary indices of autonomic arousal, interoceptive processing, and attentional engagement.

Participants first completed the Discrimination Sensitivity Task (DST) while physiological signals were recorded following a baseline measurement. Following recording, they performed the Diversity-Inclusive Hiring Task (DIHT) to assess inclusive decision-making (Figure 1). EEG, ECG, and EOG signals were recorded continuously during the DST, with heart rate, heartbeat-evoked potentials, and spontaneous blink rate derived as the primary neurophysiological indices corresponding to the study hypotheses.

60 participants took part in the study (M age = 32.26, SD = 7.28; 60.4% men, 39.6% women). Due to insufficient data quality, seven recordings were excluded from physiological analyses, resulting in a final sample of 53. Participants reported an average of 3.57 years of decision-making experience (SD = 4.41), and managed 3.74 employees (SD = 8.47), primarily in entry- to mid-level leadership roles.

Prior to the laboratory study, the video stimuli were validated in two independent online surveys to ensure reliable differentiation between conditions. In the first validation, valence $t(94) = 12.79$, $p < .001$, $d = 2.89$, and arousal $t(94) = -5.13$, $p < .001$, $d = -1.16$, differed significantly between conditions. The 30 best-performing videos from each condition were revalidated ($N = 96$), replicating significant differences in valence $t(96) = 17.46$, $p < .001$, $d = 3.86$, and arousal $t(96) = -7.88$, $p < .001$, $d = -1.74$.

EEG (32-channel 10–20 layout; Brain Products LiveAmp) and ECG/EOG (BITalino) were recorded and synchronized via the LSL with PsychoPy for stimulus presentation and markers. Prior to task onset, a three-minute baseline recording (eyes open/closed) was obtained. Analyses will be conducted at the trial level. Neural, autonomic, and ocular indices will be derived from continuous EEG, ECG, and EOG signals during the 10-second videos. These measures include blink-related (EOG) and heart rate (ECG) markers alongside EEG responses to discrimination-related stimuli.

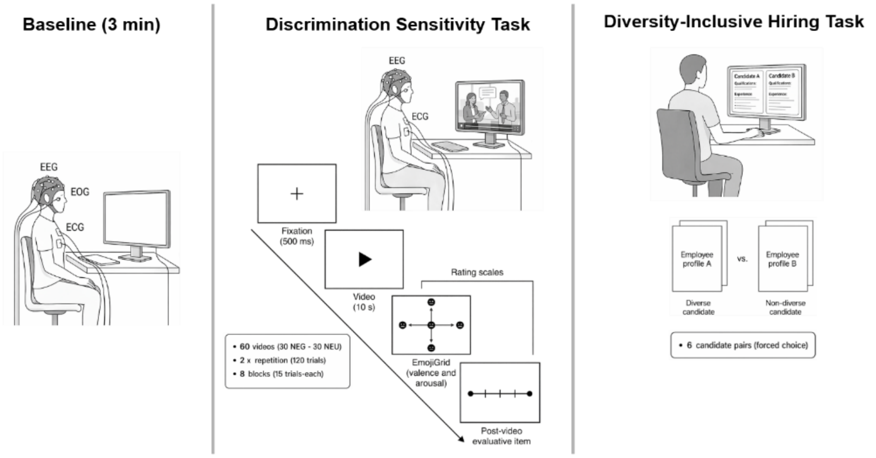


Figure 1: Overview of the experimental procedure: following a baseline recording (EEG, ECG, EOG), participants completed the DST, and then the DIHT.

Tasks

Discrimination Sensitivity Task (DST). The DST was designed to elicit affective and attentional responses to discriminatory workplace events while multimodal physiological signals were recorded. Participants viewed 10-second videos depicting workplace interactions. The stimulus set comprised 60 videos (30 discriminatory, 30 neutral), each presented twice across eight blocks (15 trials per block). Video order was randomized across blocks to minimize habituation effects and order-related biases. Each trial began with a 500ms fixation cross, followed by a 10-second video during which neurophysiological signals were recorded. After stimulus exposure, participants rated valence and arousal using the EmojiGrid [20] and completed three evaluative items assessing perceived fairness and acceptability.

Diversity-Inclusive Hiring Task (DIHT). After completing the DST and removing physiological recording devices, participants completed the DIHT. In this task, participants selected one candidate from each of six pairs of candidate profiles for participation in a leadership development program. Across randomized pairs, profiles were matched on qualifications, professional experience, performance indicators, and leadership potential. In four pairs, profiles differed in a single diversity characteristic (e.g., gender), while two pairs were fully equivalent and served as controls. Profiles were pre-validated to ensure comparable competence and suitability for the role. The DIHT operationalizes inclusive decision-making as a behavioral choice under conditions of informational equivalence. Because profiles are matched on all qualification-relevant dimensions, any systematic preference for or against candidates differing in a diversity characteristic reflects the degree to which diversity relevant information is incorporated into the selection decision. This candidate-selection paradigm provides a controlled and ecologically valid measure of hiring bias that moves beyond self-reported diversity attitudes, which are prone to social desirability inflation, by capturing actual behavioral choices. Selecting candidates from underrepresented or diversity relevant groups under conditions of matched qualifications reflects the integration of diversity relevant information into the decision process and is therefore interpreted as an inclusive decision. To quantify perceived equivalence, a comparability index was computed based on the distance of ratings from the scale midpoint (3 = equally suitable), with higher values indicating greater similarity between candidates. The index indicated high equivalence (M = 0.75), suggesting candidates were generally perceived as similarly qualified. Pairs were randomized.

Neurophysiological measures

Physiological activity was recorded continuously during the DST. EEG (32-channel 10–20 layout; Brain Products LiveAmp) was acquired at 1500 Hz, while cardiac and oculomotor signals were recorded using BITalino ECG and EOG sensors (1000 Hz). All data streams were synchronized via the Lab Streaming Layer (LSL) framework, with PsychoPy controlling stimulus presentation and sending event markers to ensure precise temporal alignment between physiological signals and experimental events.

Three neurophysiological measures were derived in direct correspondence with the study hypotheses. Heart rate (HR), extracted per video segment from R-peak intervals detected in the ECG signal, indexes autonomic arousal in response to emotionally salient stimuli [13]. Elevated HR during discriminatory relative to neutral scenarios is interpreted as heightened sympathetic activation, reflecting affective appraisal of the event. Heartbeat-evoked potentials (HEP), computed by time-locking continuous EEG activity to detected R-peaks and averaging across heartbeat epochs within each 10-second stimulus window, index interoceptive awareness and the degree to which cortical processing integrates afferent cardiac signals [12]. HEP amplitude modulation is expected to covary with the perceived salience of the discriminatory event. Spontaneous blink rate (SBR), operationalized as the number of blinks per 10-second video window detected from the EOG signal, indexes sustained attentional engagement. Reduced SBR during discriminatory scenarios reflects the involuntary suppression of blinking under heightened attentional demand, indicating deeper processing of the discriminatory content [15].

EEG preprocessing will follow established pipelines including band-pass filtering (0.1–40 Hz), re-referencing to the average reference, and artifact correction using independent component analysis. ECG data will be preprocessed using NeuroKit2, including signal cleaning, R-peak detection, and artifact correction. Recordings with more than 50% missing samples will be excluded.

Analysis Plan

Analyses will be conducted at the trial level to capture brain–body responses as they unfold during stimulus exposure. Given the within-subjects design and the nested structure of repeated observations within participants, multilevel modeling will be used as the primary analytic framework, allowing physiological responses to be examined across conditions while accounting for individual variability.

To examine condition differences in physiological responding (H1), neurophysiological measures will be compared between discriminatory and neutral scenarios, with trial order retained as a covariate to account for potential habituation effects across repeated exposures. Baseline measures will be used to control individual differences in resting physiological tone.

To examine the relationship between physiological responses and subsequent inclusive hiring decisions (H2 and H3), aggregated physiological indices from the DST will be used as predictors of decision outcomes in the DIHT, alongside subjective affect ratings. This linking approach is designed to directly test whether affective and neurophysiological responses during exposure to discriminatory scenarios translate into more inclusive behavioral choices in the subsequent task.

Subjective affective responses and neurophysiological measures will be analyzed in parallel to assess convergence across measurement modalities and evaluate the complementary contribution of physiological indices relative to self-report data alone. Unlike self-report measures, which are susceptible to social desirability and retrospective bias, physiological responses provide continuous, real-time indices of affective and attentional processes as they unfold during stimulus exposure.

Initial Results

Self-reported affective responses were assessed using the EmojiGrid, yielding standardized signed measures of valence and arousal averaged within participants for neutral (NEU) and discriminatory (NEG) workplace scenarios. Discriminatory scenarios were rated as significantly more unpleasant ($M = -0.144$, $SD = 0.061$) than neutral scenarios ($M = 0.058$, $SD = 0.047$) and elicited higher arousal (NEG: $M = 0.072$, $SD = 0.089$; NEU: $M = -0.005$, $SD = 0.070$). Paired-samples t-tests confirmed robust condition differences for both valence, $t(52) = -35.77$, $p < .001$, $d = -3.02$, and arousal, $t(52) = 6.92$, $p < .001$, $d = 0.96$. These results indicate that the video stimuli successfully differentiated between neutral and discriminatory workplace scenarios and elicited distinct affective responses during stimulus exposure.

Physiological responses were analyzed at the stimulus level to capture brain–body responses to discrimination-related events. ECG data were preprocessed using NeuroKit2, including signal cleaning, R-peak detection, and extraction of heart rate and time-domain HRV indices. Trials with fewer than five R-peaks or durations shorter than seven seconds were excluded. Mean heart rate (HR) and HRV indices (MeanNN, RMSSD) were computed per stimulus and normalized within participants. Preliminary analyses indicated a trend toward higher heart rate during discriminatory scenarios compared to neutral scenarios, $t(52) = -1.93$, $p = .060$, $d = 0.27$, accompanied by lower MeanNN, $t(52) = 1.89$, $p = .065$, $d = -0.25$. No difference was observed for RMSSD, $t(52) = -0.32$, $p = .752$. As analyses are ongoing, these findings should be considered preliminary. The observed pattern suggests modest autonomic modulation during exposure to discriminatory workplace scenarios. Further neural analyses will clarify whether brain–body responses during stimulus exposure relate to downstream inclusive decision-making in the hiring task.

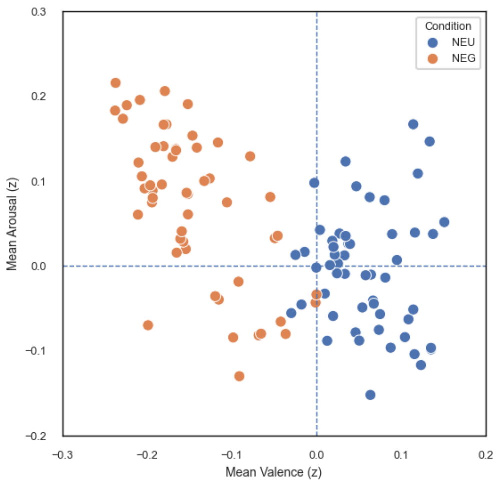


Figure 2: Mean Valence and Arousal Ratings by Condition (Neutral vs. Discriminatory Scenarios).

Discussion

The initial results confirm that the experimental paradigm successfully differentiates between neutral and discriminatory workplace scenarios. Discriminatory interactions were rated as more unpleasant and arousing than neutral scenarios, supporting their characterization as affective events within the AET framework and indicating that the video stimuli reliably evoke distinct affective responses. Preliminary autonomic analyses suggest modest physiological engagement during exposure to discriminatory content. Although these effects remain trend-level and require further validation through complete multimodal analyses, their direction is consistent with the expected increase in autonomic arousal under affective load, providing initial evidence that discriminatory workplace events may elicit measurable brain–body responses during stimulus exposure.

Situated within the broader NeuroIS literature, these preliminary findings extend prior work into a domain that has received little empirical attention. Empirical research has shown that neurophysiological signals capture cognitive and affective processes underlying decision-making in IS contexts. Foundational studies have identified neural correlates of key decision variables, such as trust in e-commerce, where Dimoka [9] linked activation in trust- and distrust-related regions to evaluative judgments in transactions. More recent work demonstrates that measures such as EEG and heart rate variability provide real-time indices of internal states and are increasingly integrated into decision-support systems. The present study applies a comparable logic to a qualitatively different domain: socially and morally charged workplace interactions. The shift from abstract economic or technological paradigms to naturalistic video-based discrimination scenarios represents a meaningful extension of the NeuroIS evidentiary base into organizational contexts where affective responses are likely to carry direct consequences for fairness-related decision outcomes.

The multimodal design, combining EEG, ECG, and EOG during video-based workplace scenarios, demonstrates the feasibility of recording concurrent physiological signals during exposure to dynamic, ecologically grounded stimuli. This represents a methodological contribution to the NeuroIS toolkit, though one that should be understood as an early step rather than a complete framework. At this stage, full multimodal integration at the analytical level has not yet been implemented. Autonomic and ocular signals are currently conceptualized as complementary sources of information that can help contextualize and situate neural responses measured with EEG, with systematic cross-signal analyses planned for subsequent work.

In subsequent analyses, ECG and EOG signals will be integrated with EEG measures to provide a more comprehensive account of autonomic and attentional dynamics during stimulus processing. This integration will allow examination of whether discriminatory workplace scenarios elicit patterns of physiological responding consistent with affective differentiation between conditions and, critically, whether those responses predict inclusive hiring behavior in the subsequent task. More broadly, the paradigm provides a basis for future research investigating how physiological responses during exposure to socially salient workplace events relate to downstream organizational decisions, an area that remains underexplored in the NeuroIS literature.

References

1. Kouchaki, M., Smith, I.H.: Moral Decision-Making in Organizations. Annual Review of Organizational Psychology and Organizational Behavior. 12, 45–72 (2025). https://doi.org/10.1146/annurev-orgpsych-110622-045715.
2. Lerner, J.S., Li, Y., Valdesolo, P., Kassam, K.S.: Emotion and Decision Making. Annu. Rev. Psychol. 66, 799–823 (2015). https://doi.org/10.1146/annurev-psych-010213-115043.
3. Todd, T.N., Lakhiwal, A., Léger, P.-M., Bala, H.: How Generative AI Review Summaries Disrupt Users' Evaluative Processes in Online Purchase Environments. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., B. Randolph, A., and R. Müller-Putz, G. (eds.)

Information Systems and Neuroscience. pp. 1–10. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-032-00815-2_1.

4. Jordan, P.J., Troth, A.C., Yan, H.: Objective and subjective measurement in applied business settings: Improving research in organizations. Australian Journal of Management. 50, 8–31 (2025). https://doi.org/10.1177/03128962241286258.
5. Teoh, Y.Y., Cunningham, W.A., Hutcherson, C.A.: Framing Subjective Emotion Reports as Dynamic Affective Decisions. Affec Sci. 4, 522–528 (2023). https://doi.org/10.1007/s42761-023-00197-y.
6. 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.: On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS Quarterly. 36, 679–702 (2012). https://doi.org/10.2307/41703481.
7. Weiss, H., Cropanzano, R.: Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. In: Research in organizational behavior: An annual series of analytical essays and critical reviews. pp. 1–74. Elsevier Science/JAI Press (1996).
8. Riedl, R., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P.A., Straub, D.W., Vom Brocke, J., Weber, B.: On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. CAIS. 27, (2010). https://doi.org/10.17705/1CAIS.02715.
9. Dimoka, A.: What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. Management Information Systems Quarterly. 34, 373–396 (2010). https://doi.org/10.2307/20721433.
10. Balconi, M., Rovelli, K., Angioletti, L., Allegretta, R.A.: Working Memory Workload When Making Complex Decisions: A Behavioral and EEG Study. Sensors. 24, 5754 (2024). https://doi.org/10.3390/s24175754.
11. Forte, G., Morelli, M., Grässler, B., Casagrande, M.: Decision making and heart rate variability: A systematic review. Applied Cognitive Psychology. 36, 100–110 (2022). https://doi.org/10.1002/acp.3901.
12. Tanaka, Y., Ito, Y., Shibata, M., Terasawa, Y., Umeda, S.: Heartbeat evoked potentials reflect interoceptive awareness during an emotional situation. Sci Rep. 15, 8072 (2025). https://doi.org/10.1038/s41598-025-92854-4.
13. Riganello, F., Vatrano, M., Tonin, P., Cerasa, A., Cortese, M.D.: Heart Rate Complexity and Autonomic Modulation Are Associated with Psychological Response Inhibition in Healthy Subjects. Entropy. 25, 152 (2023). https://doi.org/10.3390/e25010152.
14. Alyan, E., Arnau, S., Reiser, J.E., Getzmann, S., Karthaus, M., Wascher, E.: Blink-related EEG activity measures cognitive load during proactive and reactive driving. Sci Rep. 13, 19379 (2023). https://doi.org/10.1038/s41598-023-46738-0.
15. Maffei, A., Angrilli, A.: Spontaneous blink rate as an index of attention and emotion during film clips viewing. Physiology & Behavior. 204, 256–263 (2019). https://doi.org/10.1016/j.physbeh.2019.02.037.
16. Storbeck, J., Clore, G.L.: Affective Arousal as Information: How Affective Arousal Influences Judgments, Learning, and Memory. Social & Personality Psych. 2, 1824–1843 (2008). https://doi.org/10.1111/j.1751-9004.2008.00138.x.

17. Zheng, Y., Yang, Z., Jin, C., Qi, Y., Liu, X.: The Influence of Emotion on Fairness-Related Decision Making: A Critical Review of Theories and Evidence. Front. Psychol. 8, 1592 (2017). https://doi.org/10.3389/fpsyg.2017.01592.
18. Li, Q., Lai, X., Li, T., Madsen, K.H., Xiao, J., Hu, K., Feng, C., Fu, D., Liu, X.: Brain responses to self- and other- unfairness under resource distribution context: Meta-analysis of fMRI studies. NeuroImage. 297, 120707 (2024). https://doi.org/10.1016/j.neuroimage.2024.120707.
19. Robinson, M.D., Klein, R.J., Irvin, R.L., McGregor, A.Z.: Attention to emotion and reliance on feelings in decision-making: Variations on a pleasure principle. Cognition. 217, 104904 (2021). https://doi.org/10.1016/j.cognition.2021.104904.
20. Toet, A., Kaneko, D., Ushiama, S., Hoving, S., De Kruijf, I., Brouwer, A.-M., Kallen, V., Van Erp, J.B.F.: EmojiGrid: A 2D Pictorial Scale for the Assessment of Food Elicited Emotions. Front. Psychol. 9, 2396 (2018). https://doi.org/10.3389/fpsyg.2018.02396.

Towards Mental Fatigue-Adaptive Systems: A Pilot Interview Study

Lukas Schick[1], Luca L. Bennardo[1], Chiara Krisam[1], Jella Pfeiffer[1], Christof Weinhardt[1], Michael T. Knierim[1,2]

[1]Institute for Information Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany
{lukas.schick, luca.bennardo, chiara.krisam, pfeiffer, weinhardt, michael.knierim}@kit.edu
[2]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

Abstract. Mental fatigue is increasingly common in digital environments. At the same time, wearable EEG devices embedded in headphones enable real-time detection of mental fatigue, creating new opportunities for adaptive systems. However, user-centered requirements for biosignal-based fatigue-adaptive systems remain largely unexplored. To address this gap and derive design implications, we conducted 15 laddering interviews with knowledge workers across three scenarios: individual work, remote meetings, and online shopping. Results show that users value fatigue mitigation, decision support, privacy, and autonomy. Effective systems should provide context-sensitive suggestions, unobtrusive notifications, and transparent data handling to support performance and well-being.

Keywords: Mental Fatigue, Adaptive System, EEG, Wearables, Interview Study

Introduction

As information technology has become ubiquitous, the demands it poses, such as greater availability of information and constant reachability [1–3], have put additional strain on our cognitive resources. As a result, considerable parts of the working population report suffering from mental fatigue [4, 5]. Mental fatigue occurs after prolonged periods of engaging with cognitively demanding tasks. It is characterized through decreased performance, lapses in concentration, and reduced mental well-being [4]. For professionals in safety-critical areas like healthcare personnel, this puts human safety at risk [6], whereas for knowledge workers more generally, economic damage and long-term health risks like burnout are possible consequences [7]. Outside of work, mental fatigue can also affect daily life, e.g., by negatively impacting physical performance [8] and the ability to participate in social activities [9].

Mental fatigue can be passively assessed with neurophysiological data, electro-encephalography (EEG) being the most established method [10]. When combined with suitable machine learning algorithms, this allows for real-time insights [11]. Together with advances in wearable sensing devices, which facilitate the setup and use of EEG compared to traditional, clinical systems [12, 13], this opens the door for biosignal-adaptive systems [14] which can be used in real-world settings. Concretely, headphone

EEG has shown potential for wearable cognitive load [15], flow [16] and mental fatigue [17] detection. However, user-centric design requirements for EEG-based fatigue-adaptive systems remain largely unexplored. As pointed out by Kögel et al. [18], the vast majority of brain-computer-interface-related studies either focus on the technical aspects of these systems or use quantitative data to assess the user perspective. Qualitative studies addressing the needs and expectations of potential users in knowledge work settings are lacking, with the few existing studies limited to the clinical domain [19, 20]. As such EEG-based systems will play a prominent role in the user's work and leisure activities, potentially influencing their behaviour and prompting self-reflection and interoception, it is important to consider the user perspective when designing these systems to avoid friction and promote system acceptance. Furthermore, even though several mental fatigue mitigation strategies have been proposed [21] which mainly function through cognitive load reduction [22] and motivation increases [23], these strategies have thus far not been linked to biosignal-adaptive systems. Which interventions such systems should perform, how, and in which context, has so far not been studied. Thus, we pose the following research question:

RQ: How to design mental fatigue-adaptive systems based on headphone EEG data across knowledge work and leisure contexts?

To answer this question, we conducted 15 laddering interviews across three different contexts – individual knowledge work, remote meetings, and shopping under pressure – deriving design implications and proposing how such a system could function in different contexts.

Conceptual Foundations

In EEG data, the power of the theta (4-8 Hz) and alpha (8-13 Hz) wavebands and their ratios are seen as reliable markers of mental fatigue [10, 24]. As these frequency ranges are associated with an awake yet unfocused brain state [25], rising theta and alpha power during cognitively demanding tasks indicates increasing mental fatigue [24]**.** Importantly, global changes in these bands, rather than region-specific effects alone, are sufficient for detection [10], enabling mental fatigue measurement even with low spatial coverage setups such as headphone EEG.

Recently, there have been several publications in the NeuroIS community concerned with biosignal-adaptive systems that trigger adaptations based on the user's cognitive state and behavior, thus offering timely and situational support. Systems capable of adapting the responses of genAI assistants [26] and suggesting breaks in remote meetings [27] based on cognitive load, assessed through EEG and eye tracking, have been proposed. Further proposed systems include gaze-based trust building in remote meetings [28], and visual complexity adjustments in Virtual Reality environments, where EEG is used to assess attention [29]. Most pertinently, an EEG-based mental fatigue alert system for drone operators has been developed [30]. However, the aim of this system is increased performance rather than fatigue mitigation. Furthermore, the 64-channel EEG setup which was used is unsuited for everyday scenarios.

So, even though mental fatigue detection in everyday situations is possible through headphone EEG and cognitive state-adaptive systems are being developed, a user-led study on how a mental fatigue-adaptive system should be designed is lacking.

Methodology

Laddering Interviews. We conducted laddering interviews to elicit Attribute–Consequence–Value (ACV) chains by repeatedly asking "why" questions to uncover underlying user values [31]. Laddering is commonly used to guide the design process for information systems [32]. To ensure scalability and structure, we used LadderChat [33], a conversational agent that guides participants through hard laddering sequences with interactive probing. Participants completed three separate chat sessions, one per scenario, allowing systematic comparison of consequences and values while maintaining contextual separation. Participants were sent a link to the study and completed it remotely, with interviews lasting 39 minutes on average. Before the interview started, participants were informed about the purpose of the study, as well as the laddering interview technique, and gave their consent. At the start of the interview, participants were confronted with short descriptions for each scenario and were asked to rank the scenarios in order of importance to them. This ranking would then dictate the order of scenarios for the interview. The placeholder order was always: individual work, remote meeting, online shopping.

Scenarios. Two knowledge work settings (mentally demanding individual tasks; large, remote meetings without active participation) and one leisure setting (purchasing an expensive product online under time pressure) were considered.[3] Prior research shows that cognitively demanding office work and technology-mediated collaboration induce mental fatigue and impair performance and well-being [24, 34–37]. Although less studied, time pressure and information overload have also been shown to negatively affect decision-making in shopping [38, 39]. Including three distinct scenarios from work and leisure settings serves multiple purposes. Although all three contexts involve mental fatigue, they differ in terms of task type (e.g., open-ended cognitive work vs. structured decision-making), social embeddedness (e.g., individual vs. collaborative), and duration of fatigue exposure (e.g., sustained effort over varying time spans vs. short-term pressured choices). By spanning work and leisure domains, we aim to identify design implications that are universal to fatigue-adaptive systems, as well as those that are highly context-dependent. In all three scenarios, participants imagined wearing EEG headsets that detect mental fatigue while performing activities. Participants were asked to describe the desired system functionality and underlying reasons for their choices in each scenario. When prompted about system requirements in the different scenarios, participants were not limited in the amount of ideas they could contribute, even if the ideas were very different from one another. Therefore, the ideas did not need to build on one another or be implementable simultaneously.

Participants. 15 knowledge workers (9 male, 6 female; age 20–35, M = 27) participated in the study. The majority held a Master's degree (n = 11); others held a Bachelor's degree (n = 2) or completed secondary education (n = 2). Professions included

[3] The scenario descriptions, an exemplary chat excerpt and an ACV tree are uploaded to: https://osf.io/xte5u/overview?view_only=47a96443b4834ee98c44e13dc7605626

researchers (5), IT professionals (4), consultants (3), and student assistants (3). Participants were recruited through convenience sampling and include researchers and student assistants from the authors' institutions. No monetary incentive for participation was provided. Table 1 gives an overview of the participants' demographic information, along with their stated scenario priorities.

Table 1. Participant preferences for scenarios along with demographic information.

ID	Prio 1	Prio 2	Prio 3	Age	Gender	Occupation
P1	I*	M**	S***	30	m	Researcher
P2	I	M	S	26	m	Researcher
P3	M	I	S	27	m	IT
P4	I	M	S	28	m	Researcher
P5	I	M	S	27	f	Researcher
P6	I	S	M	31	m	Research Assistant
P7	I	S	M	22	f	Research Assistant
P8	I	S	M	27	f	Research Assistant
P9	M	S	I	35	f	Researcher
P10	I	M	S	26	f	IT
P11	I	M	S	27	m	Consultant
P12	S	I	M	24	m	IT
P13	S	I	M	20	m	Consultant
P14	I	M	S	26	f	IT
P15	S	M	I	22	m	Consultant

* I = Individual Work, **M = Meeting, ***S = Shopping (Placeholder order was I-M-S)

Analysis. After manually verifying LadderBot's automated matrices and ACV chains, three researchers iteratively coded the data to derive design implications. First, a codebook was developed from three joint analyses. The remaining interviews were then independently coded by scenario. Themes were regularly refined to ensure consistency.

Mental Fatigue-Adaptive Systems for Work and Life

Utilizing the ACV model inherent to laddering interviews, we first discuss the attributes for the different scenarios mentioned by interviewees (see Table 2 for an overview). Afterwards, the underlying benefits these attributes provide and which goals they satisfy – as extracted from the consequences and values – are considered (see Table 3 for an overview). Finally, we derive design implications and propose how the system should function in different contexts. It should be noted that multiple attributes, consequences and values were extracted per interview, meaning that multiple, sometimes contradictory views could be expressed by each participant.

Attributes

Individual Work. The most frequently mentioned attribute was a *break suggestion* (P1, P2, P4-P7, P9-P12, P14, P15), including recommendations for both break length (P7) and activity (P14), expressing a strong desire for immediate mental fatigue reduction. In contrast, some participants preferred maintaining workflow by *blocking interruptions* instead of pausing to improve focus (P2, P8, P10). Others suggested recommending less demanding tasks (P1, P14) or agentic AI support for task completion (P11). *Interoceptive features* such as visualizing fatigue history (P12) and productivity comparisons over time (P9, P12) were also mentioned. Regarding implementation, participants emphasized *low visual prominence* and *infrequent notifications* (P1, P3, P5, P6, P8), *user control* over interventions (P7), and *privacy-preserving design* (P5, P13).

Remote Meetings. As in the previous scenario, the most frequently mentioned attribute was a *break suggestion* (P3, P5, P9, P10), proposing short pauses with concrete recovery guidance. Three participants (P1, P5, P11) emphasized *early warnings* so the system would intervene before fatigue negatively affects them. Regarding implementation, three participants preferred *discrete notifications* (P5, P9, P11), whereas P7 asked for an *obvious and short alert*. Three participants (P4, P14, P15) suggested basing the mechanism on the *fatigue states of all participants*, with P14 and P15 proposing *group-level warnings*. *Privacy concerns* were raised by P4 and P14, and P4 requested that the system explicitly *displays the current fatigue state*. As alternatives to interruption, two participants (P8, P12) proposed automatically generated *transcripts or summaries,* while P7 suggested *keyword detection* to notify users when relevant topics arise. Finally, P13 proposed *automatically logging off* and notifying the team.

Shopping under Pressure. In the shopping context, most participants wanted the system to *provide relevant product information*, particularly technical details, prices, reviews, and delivery times (P4, P7, P8, P10, P12). The second most-mentioned attribute was *getting competent advice* (P11, P13-P15) or guidance in general to speed up the process (P9, P13, P15). Three participants specifically asked the system to *detect their preferences and provide personalized recommendations* (P3, P7, P13). Three other participants wanted the system to *show a selection of relevant products*, potentially with underlying product comparisons (P7, P10, P12). More simple measures included playing calming *background music* (P15) or providing a *warning about the mental fatigue state* (P2). With regards to how the adaptations should be designed, it was mentioned that mental fatigue should *not be abused* to encourage impulsive buying but rather take measures to reduce it (P1, P4, P8). The system should act *unbiased* (P11) and be implemented as an *open-source* system (P1) to ensure that the system transparently helps users, rather than following revenue-maximizing goals.

Table 2. Overview of identified attributes per scenario.

Individual Work	Remote Meetings	Shopping Under Pressure
Break suggestion		-
Discrete notifications		-
Privacy-preserving design		-
-	Early warnings	
Block interruptions	Short, prominent notifications	Provide product information
Fatigue-adapted task recommendation	Meeting-wide fatigue mechanism	Give personalized advice
AI task support	Group-level fatigue warnings	Find selection of relevant products
Show fatigue history	Generate meeting summaries	Calming background music
Show productivity level	Display current fatigue state	Don't encourage impulse buying
User control over system	Keyword detection for relevance notification	Unbiased system
-	Automatically log off from meeting	Open-source system

Consequences & Values

Across all three scenarios, participants consistently linked fatigue detection and reduction to improved performance and well-being (e.g., P1, P13). Fatigue reduction was broadly associated with higher productivity and better decisions (e.g., P2, P7, P15), improved mood and reduced stress (e.g., P2, P11, P14), and broader mental health benefits (e.g., P3, P12, P13). Moreover, autonomy-related values emerged across contexts: participants emphasized retaining freedom of action (P5, P6), decision autonomy (P4), and increased control and confidence (P8, P9, P10, P14, P15). Trust in the system and concerns that it should not become burdensome were also raised (e.g., P4, P6, P11). At the same time, clear contextual differences appeared. In individual work, self-awareness and longer-term flourishing, such as improved well-being and living a happy life (e.g., P4, P11, P13), were emphasized. In remote meetings, fatigue was primarily connected to social functioning, including being an active, reliable team member and avoiding social awkwardness (e.g., P7, P9, P10, P11, P15). In shopping, consequences were more utility-oriented, emphasizing time value, financial benefits, and long-term satisfaction (e.g., P1, P7, P12, P14, P15).

Table 3. Overview of identified consequences and values per scenario.

Individual Work	Remote Meetings	Shopping Under Pressure
	Improved performance	
	Improved well-being	
	Higher productivity	
	Improved decision making	
	Improved mood	
	Reduced stress	
	Retain freedom of action	
	Retain decision autonomy	
	Improve confidence in actions	
	Trust in system	
	No burden through system	
Increase self-awareness	Being a reliable team member	Being time efficient
Long-term personal growth	Avoiding social awkwardness	Financial benefits and purchase satisfaction

Design Implications and System Functionality

Deriving design implications from the ACV chains analysed above, we propose how a mental-fatigue adaptive system should function in different contexts. Concerns that the system itself should not be burdensome support the planned use of EEG headphones to collect data discreetly and continuously. The fatigue state is then determined through user-specific machine learning models with relative power of pertinent frequency bands and their ratios as input. If mental fatigue is detected, the system adapts contextually.

In work settings, the system should primarily suggest breaks, tailoring timing, duration, and activity to fatigue level and task demands. During individual work, it should additionally block irrelevant interruptions and offer task-related support (e.g., agentic AI assistance). In remote meetings, fatigue states of all participants – if available – should inform coordinated interventions, while individual support may include automated summaries or keyword-based alerts. In the shopping scenario, in which participants were presented with an explicit decision-making task under time pressure, the most desirable category of adaptation measures was found to be more efficient and effective goal facilitation, which involves providing information, personalized recommendations, and decision support to reduce cognitive demands and indirectly mitigate fatigue. However, this finding likely reflects the scenario's framing of decision-making and should not be considered a general principle for all leisure contexts. In any case, our findings clearly suggest that although the values and motivations for using such systems are similar across all three scenarios, the desired measures depend heavily on the fatigue-inducing situation and differ more strongly between related and unrelated activities (e.g., work vs. leisure).

Across contexts, notifications should be clear yet unobtrusive, favoring false negatives over false positives. Interventions must remain suggestions to preserve user autonomy. Data privacy must be ensured, with no sharing to external parties.

Discussion and Outlook

Our findings indicate that users perceive mental fatigue-adaptive systems as helpful for supporting their work and everyday life activities. However, participants emphasized that interventions should remain suggestions rather than enforced actions. Prior research has already pointed to the importance of maintaining user control, for example by allowing users to override system behavior or by framing adaptations as optional rather than enforced [27, 28]. Similarly, concerns regarding privacy and transparency, also emphasized in prior work [27, 28], were prominent in our findings, with participants stressing the need for responsible data handling and control over personal information. Our results further highlight the importance of context in shaping how support should be operationalized, extending prior work that typically focuses on one context [30]. In work settings, participants preferred fatigue mitigation, e.g. through break recommendations, whereas in the shopping scenario users valued decision support. These differences highlight the importance of designing adaptive systems with careful consideration of the specific context. At the same time, general user values remained largely consistent across settings, like the goals of performance, well-being, and maintaining a sense of control.

This study contributes to research and practice in two ways. First, it provides empirical insights into user requirements for mental fatigue-adaptive systems across multiple contexts, extending prior work that has largely focused on single scenarios. By showing that user preferences vary not only between individuals but also across contexts, our findings highlight the need for future research to explicitly account for contextual variability when studying adaptive systems. Second, the study derives concrete design implications for practice, emphasizing that effective systems should provide context-sensitive support while preserving user autonomy and ensuring transparent, privacy-preserving data handling. Together, these insights offer a foundation for designing and evaluating user-centered fatigue-adaptive systems in real-world settings.

The proposed system functionality offers a foundation for developing and empirically testing such systems in real-world settings. Future research can build on this by evaluating the identified design implications in larger and more diverse samples and by further analyzing the detailed ACV chains. Overall, this paper provides an initial overview of shared and context-specific patterns and outlines promising directions for the design of mental fatigue-adaptive systems.

Acknowledgements Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – GRK2739/2 – Project Nr. 447089431 – Research Training Group: KD2School – Designing Biosignal-Adaptive Systems for Decision-Making Processes

References

1. Riedl, R.: On the biology of technostress. ACM SIGMIS Database DATABASE Adv. Inf. Syst. 44, 18–55 (2012). https://doi.org/10.1145/2436239.2436242.
2. Ayyagari, R., Grover, V., Purvis, R.: Technostress: Technological Antecedents and Implications1. MIS Q. 35, 831-A10 (2011). https://doi.org/10.2307/41409963.
3. Sanjeeva Kumar, P.: TECHNOSTRESS: A comprehensive literature review on dimensions, impacts, and management strategies. Comput. Hum. Behav. Rep. 16, 100475 (2024). https://doi.org/10.1016/j.chbr.2024.100475.
4. Boksem, M.A.S., Tops, M.: Mental fatigue: Costs and benefits. Brain Res. Rev. 59, 125–139 (2008). https://doi.org/10.1016/j.brainresrev.2008.07.001.
5. Tanaka, M., Ishii, A., Watanabe, Y.: Neural effects of mental fatigue caused by continuous attention load: A magnetoencephalography study. Brain Res. 1561, 60–66 (2014). https://doi.org/10.1016/j.brainres.2014.03.009.
6. Kunasegaran, K., Ismail, A.M.H., Ramasamy, S., Gnanou, J.V., Caszo, B.A., Chen, P.L.: Understanding mental fatigue and its detection: a comparative analysis of assessments and tools. PeerJ. 11, e15744 (2023). https://doi.org/10.7717/peerj.15744.
7. Yaacob, H., Hossain, F., Shari, S., Khare, S.K., Ooi, C.P., Acharya, U.R.: Application of Artificial Intelligence Techniques for Brain–Computer Interface in Mental Fatigue Detection: A Systematic Review (2011–2022). IEEE Access. 11, 74736–74758 (2023). https://doi.org/10.1109/ACCESS.2023.3296382.
8. Van Cutsem, J., Marcora, S., De Pauw, K., Bailey, S., Meeusen, R., Roelands, B.: The Effects of Mental Fatigue on Physical Performance: A Systematic Review. Sports Med. 47, 1569–1588 (2017). https://doi.org/10.1007/s40279-016-0672-0.
9. Ream, E., Richardson, A.: Fatigue: a concept analysis. Int. J. Nurs. Stud. 33, 519–529 (1996). https://doi.org/10.1016/0020-7489(96)00004-1.
10. Tran, Y., Craig, A., Craig, R., Chai, R., Nguyen, H.: The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. Psychophysiology. 57, e13554 (2020). https://doi.org/10.1111/psyp.13554.
11. Monteiro, T.G., Skourup, C., Zhang, H.: Using EEG for Mental Fatigue Assessment: A Comprehensive Look Into the Current State of the Art. IEEE Trans. Hum.-Mach. Syst. 49, 599–610 (2019). https://doi.org/10.1109/THMS.2019.2938156.
12. Bleichner, M.G., Debener, S.: Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. Front. Hum. Neurosci. 11, (2017). https://doi.org/10.3389/fnhum.2017.00163.
13. Niso, G., Romero, E., Moreau, J.T., Araujo, A., Krol, L.R.: Wireless EEG: A survey of systems and studies. NeuroImage. 269, 119774 (2023). https://doi.org/10.1016/j.neuroimage.2022.119774.

14. Schultz, T., Maedche, A.: Biosignals meet Adaptive Systems. SN Appl. Sci. 5, 234 (2023). https://doi.org/10.1007/s42452-023-05412-w.
15. Knierim, M.T., Zimny, C., Ivucic, G., Röddiger, T.: Advancing Wearable BCI: Headphone EEG for Cognitive Load Detection in Lab and Field. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 9, 1–26 (2025). https://doi.org/10.1145/3712283.
16. Knierim, M.T., Nadj, M., Hariharan, A., Weinhardt, C.: Flow Neurophysiology in Knowledge Work: Electroencephalographic Observations from Two Cognitive Tasks: In: Proceedings of the 5th International Conference on Physiological Computing Systems. pp. 42–53. SCITEPRESS - Science and Technology Publications, Seville, Spain (2018). https://doi.org/10.5220/0006926700420053.
17. Schick, L., Frey, E., Putze, F., Knierim, M.T.: Rest Assured: Detecting Mental Fatigue and Recovery with EEG Headphones. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–17. ACM, Barcelona, Spain (2026). https://doi.org/10.1145/3772318.3791781.
18. Kögel, J., Schmid, J.R., Jox, R.J., Friedrich, O.: Using brain-computer interfaces: a scoping review of studies employing social research methods. BMC Med. Ethics. 20, 18 (2019). https://doi.org/10.1186/s12910-019-0354-1.
19. Schicktanz, S., Amelung, T., Rieger, J.W.: Qualitative assessment of patients' attitudes and expectations toward BCIs and implications for future technology development. Front. Syst. Neurosci. 9, (2015). https://doi.org/10.3389/fnsys.2015.00064.
20. Klein, E., Goering, S., Gagne, J., Shea, C.V., Franklin, R., Zorowitz, S., Dougherty, D.D., Widge, A.S.: Brain-computer interface-based control of closed-loop brain stimulation: attitudes and ethical considerations. Brain-Comput. Interfaces. 3, 140–148 (2016). https://doi.org/10.1080/2326263X.2016.1207497.
21. Proost, M., Habay, J., De Wachter, J., De Pauw, K., Rattray, B., Meeusen, R., Roelands, B., Van Cutsem, J.: How to Tackle Mental Fatigue: A Systematic Review of Potential Countermeasures and Their Underlying Mechanisms. Sports Med. 52, 2129–2158 (2022). https://doi.org/10.1007/s40279-022-01678-z.
22. Mizuno, K., Tanaka, M., Yamaguti, K., Kajimoto, O., Kuratsune, H., Watanabe, Y.: Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behav. Brain Funct. 7, 17 (2011). https://doi.org/10.1186/1744-9081-7-17.
23. Dora, J., Van Hooff, M.L.M., Geurts, S.A.E., Kompier, M.A.J., Bijleveld, E.: The effect of opportunity costs on mental fatigue in labor/leisure trade-offs. J. Exp. Psychol. Gen. 151, 695–710 (2022). https://doi.org/10.1037/xge0001095.
24. Riedl, R., Kostoglou, K., Wriessnegger, S.C., Müller-Putz, G.R.: Videoconference fatigue from a neurophysiological perspective: experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). Sci. Rep. 13, 18371 (2023). https://doi.org/10.1038/s41598-023-45374-y.
25. Müller-Putz, G.R., Riedl, R., Wriessnegger, S.C.: Electroencephalography (EEG) as a Research Tool in the Information Systems Discipline: Foundations, Measurement, and Applications. Commun. Assoc. Inf. Syst. 37, (2015). https://doi.org/10.17705/1CAIS.03746.

26. Schulz, T., Krisam, C., Seitz, J.: EyeGPT: A Cognitive Load-Adaptive GenAI Assistant with Eye Tracking for Programming Education. NeuroIS 2025 Proc.
27. Krisam, C., Seitz, J., Knierim, M.T., Maedche, A.: A Cognitive Load-Adaptive Microbreak Intervention in Video Meeting Systems: First Results from a DSR Project. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience. pp. 327–335. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9_28.
28. Reuscher, T.F., Seitz, J., Greif-Winzrieth, A., Langner, M., Soballa, K., Maedche, A., Nieken, P.: Designing Gaze-Adaptive Visualizations of Mutual Gaze for Video Meetings. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience. pp. 247–255. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9_21.
29. Chiossi, F., Ou, C., Gerhardt, C., Putze, F., Mayer, S.: Designing and Evaluating an Adaptive Virtual Reality System using EEG Frequencies to Balance Internal and External Attention States. Int. J. Hum.-Comput. Stud. 196, 103433 (2025). https://doi.org/10.1016/j.ijhcs.2024.103433.
30. Hinss, M.F., Brock, A.M., Roy, R.N.: EEG-based Adaptive Visual Alerts to Mitigate UAS Operator Mental Fatigue: A Technical Proof of Concept. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. pp. 1–7. ACM, Yokohama Japan (2025). https://doi.org/10.1145/3706599.3719921.
31. Rietz, T., Maedche, A.: Ladderbot—A conversational agent for human-like online laddering interviews. Int. J. Hum.-Comput. Stud. 171, 102969 (2023). https://doi.org/10.1016/j.ijhcs.2022.102969.
32. Rietz, T., Maedche, A.: LadderBot: A Requirements Self-Elicitation System. In: 2019 IEEE 27th International Requirements Engineering Conference (RE). pp. 357–362. IEEE, Jeju Island, Korea (South) (2019). https://doi.org/10.1109/RE.2019.00045.
33. Hanschmann, L., Mokelke, M., Maedche, A.: LadderChat An LLM-Based Conversational Agent for Laddering Interviews. In: Chatbots and Human-Centered AI: 8th International Workshop, CONVERSATIONS 2024, Thessaloniki, Greece, December 4–5, 2024, Revised Selected Papers. pp. 48–65. Springer-Verlag, Berlin, Heidelberg (2025). https://doi.org/10.1007/978-3-031-88045-2_4.
34. Nesher Shoshan, H., Wehrt, W.: Understanding “Zoom fatigue”: A mixed-method approach. Appl. Psychol. 71, 827–852 (2022). https://doi.org/10.1111/apps.12360.
35. Döring, N., Moor, K.D., Fiedler, M., Schoenenberg, K., Raake, A.: Videoconference Fatigue: A Conceptual Analysis. Int. J. Environ. Res. Public. Health. 19, 2061 (2022). https://doi.org/10.3390/ijerph19042061.
36. Brazaitis, M., Satas, A.: Regular short-duration breaks do not prevent mental fatigue and decline in cognitive efficiency in healthy young men during an office-like simulated mental working day: An EEG study. Int. J. Psychophysiol. 188, 33–46 (2023). https://doi.org/10.1016/j.ijpsycho.2023.03.007.

37. Lorist, M.M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., Meijman, T.F.: Mental fatigue and task control: Planning and preparation. Psychophysiology. 37, 614–625 (2000). https://doi.org/10.1111/1469-8986.3750614.
38. Chen, Y.-C., Shang, R.-A., Kao, C.-Y.: The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. Electron. Commer. Res. Appl. 8, 48–58 (2009). https://doi.org/10.1016/j.elerap.2008.09.001.
39. Amirpur, M., Benlian, A.: Buying under Pressure: Purchase Pressure Cues and their Effects on Online Buying Decisions. ICIS 2015 Proc. (2015).

Framing Bias and Financial Auditor Behavior: An Eye-Tracking Perspective

Prince Teye[1, 2]

[1] UCLy (Lyon Catholic University), ESDES, Lyon, France.
UCLy (Lyon Catholic University), UR CONFLUENCE : Sciences et Humanités (EA1598), Lyon, France
pteye@univ-catholyon.fr
[2] iaelyon, Magellan (EA3713), Jean Moulin University, Lyon, France
prince.teye@univ-lyon3.fr

Abstract. This study investigates the impact of the framing bias on financial auditor behavior during the identification and diagnosis of audit evidence indicative of aggressive financial reporting. Specifically, we investigate the visual behaviors of susceptible auditors vs. non-susceptible auditors during a simple audit task. We test our predictions using eye-tracking, in a controlled experiment where participants are tasked with performing an audit of financial reporting in an Audit Management Information System (AMIS). Results show that susceptible auditors exhibit higher total duration of fixations associated with ineffective detection for target information. By identifying the visual attention differences of susceptible and non-susceptible auditors, we elucidate how the framing bias may influence auditor performance.

Keywords: Eye-tracking·Financial Auditing·Framing Bias

Introduction

Audit management information systems (AMIS) are essential to the performance of audits [1, 2]. In spite of the increased awareness of the detrimental effects of cognitive biases on auditors' decisions, recent studies show that some auditors still remain susceptible to them [3]. In this study, we seek to elucidate the visual behavior of auditors susceptible to the framing bias in the performance of a simple audit task in an AMIS. The framing bias has been identified to be an important bias in auditing due to its possibility of leading to an increase in the risk of error regarding ongoing judgment [4]. An accurate understanding of the visual behavior associated with the framing bias could eventually help design better choice environments in an AMIS by attracting auditors' attention to salient items in a predictable way by means of nudges [1].

A framing bias occurs when a modification in the description of a task, which does not alter its normative meaning, changes the decision that is made [5]. The framing effect is thus characterized by inconsistencies in decisions across tasks which remain fundamentally unchanged. In auditing, studies carried out concerning the framing bias, primarily focus on its existence, effects, or debiasing techniques [6]. Again, cognitive biases in auditing are generally studied in the context of complex tasks since these tasks are associated with a higher risk of error. Consequently, decisions related to simple audit tasks remain largely unexplored [7]. From a practical standpoint, investigating simple audit tasks is necessary since novice auditors begin their practice with such tasks. Furthermore, an audit mission is generally composed of both simple and complex tasks making it impractical to overlook simple tasks.

The main objective of this study is to explore the differences in the attentional characteristics associated with the framing bias in a simple aggressive reporting detection task. In order to highlight this difference, we use eye-tracking technology. Aggressive financial reporting is the use of optimistic projections in the accounting standards to create financial statements that present a more positive outlook of a company than is actually the case. These actions are taken to give the investment community a falsely enhanced view of a business, or for the personal gain of management [1].

Our research offers two major contributions. From a theoretical standpoint, this study fills the gap of understanding the psychological construct associated with biased auditors behavior during the identification and diagnosis of audit evidence indicative of aggressive financial reporting. With the rise of remote work spurred by the recent COVID-19 pandemic, audits have largely shifted from mostly hardcopy materials to digital trails. Thus, the importance of understanding the traits of a skeptical auditor as per his attention measured with his eye movements on digital platforms cannot be overstated. In doing so, this paper responds to the call by Lynch and Andiola [8] for the application of eye-tracking in accounting and auditing research. Second, from a practical standpoint, this paper underscores the need, in the pursuit of high quality audits, to understand the cognitive profiles of individual auditors. Audit managers and seniors should be guided by the effect of the framing bias on novice professionals and collaborate with UX developers to elaborate AMIS interfaces best suited for different profile categories.

Prior Research and Hypotheses Development

The study of cognitive biases originates from the field of psychology [5, 9]. In adapting these studies to the auditing domain, Shanteau [10] delineates three approaches: replication studies (accurate reproduction of the original studies using auditors as subjects), adaptation studies (spin-offs from the original studies but concepts modified to reflect accounting/auditing issues), and problem-driven studies (exclusively focused on accounting/auditing issues, differing methodologically from original studies and not considered as spin-offs). In auditing, most papers relating to cognitive biases focus on adaptation studies, and problem driven studies hence not much is known from the replication standpoint [11, 12]. In view of this gap in the literature, we employ a replication approach of the original studies of cognitive biases in order to identify the level of susceptibility of auditors. Furthermore, an advantage of a replication approach is that it is more faithful to the original concept of the framing bias and has a wider applicability. Again, the replication study adopted in this research is better suited to a more enduring trait form of the bias compared to a temporary situational form which is reflected in the majority of adaptation studies [4, 6]. Consequently, we formulate the following hypothesis:

H1 Auditors are susceptible to the framing bias

Eye-tracking literature indicate that, on average, ineffective searches of target information are correlated with a higher number of fixations of the stimulus, and that erroneous detections of target information are associated with longer and more frequent fixations [13, 14]. Again, the possibility for successful detection decreases with time, which could be due to cognitive resource depletion through stimulus encoding processes [15, 1]. Auditing research suggests that the framing bias is associated with poorer risk assessments and inefficient hypothesis activations [4, 16, 6]. Given that the framing bias leads to poorer outcomes in auditing, we argue that in eye-tracking terms, it may lead to ineffective searches for target information characterized by a higher number of fixations. This may further lead to a depletion of cognitive resources. Consequently, we hypothesize the following:

H2 The framing bias is associated with a higher total duration of fixations

H3 The framing bias leads to a poorer detection of aggressive financial reporting items

Research Method

We conducted an experiment with 40 novice professional financial auditors with work experience ranging from 3 months to 1 year. The use of novices for this study is justified as a result of various studies indicating a higher susceptibility of novices to cognitive biases compared to their more experienced counterparts [12, 17]. Furthermore, novices are more likely to perform simple tasks. The experimental design was approved by the Institutional Review Board.

Research Design and Protocol

In order to confirm our hypotheses, we conducted a computerized test in which we tasked our participants with examining pieces of audit evidence adapted from Phillips [18, 19]. Participants undertook the test in the laboratory; they first had to read and accept the terms and conditions of participating. Subsequently, participants read the instructions for the audit exercise, which required them to undertake a self-paced review of audit evidence about a fictitious company. The instructions for the audit exercise were preceded by background information about the company and key information about the audit, such as the level of materiality and the accounting year.

Consequently, each participant had to inspect, in the fictitious interface of an AMIS, 6 pieces of audit evidence emanating from 6 distinct accounts in the financial statements. The order of appearance of the audit evidence was fully randomized. Of the 6 items, 2 were aggressive. After having examined all the 6 items of the audit evidence at their own pace, participants manually moved to the next page where they performed a free recall task which involved the identification of the audit items adjudged aggressive. Given that this research is focused on simple, low mental effort tasks, we take steps to control the level of difficulty. As recommended by Brosnan et al [20], we take the following steps: we remove superfluous survey questions, we use short sentences and simple language, and we avoid technical terms.

After that, the Hurtt professional skepticism scale (HPSS) was administered. The Hurtt scale measures ex ante an individual's level of trait professional skepticism [21]. Professional skepticism is an attitude that includes a questioning mind, being alert to conditions which may indicate possible misstatement due to error or fraud, and a critical assessment of evidence [19, 21]. Following this, participants took the test relating to the framing bias.

The material for the framing bias was obtained from Tversky and Kahneman [22] on the framing of acts. This task manipulates risk aversion by presenting two frames which indicate two decisions to be made. The first decision point is framed as a riskless prospect whereas the second decision is framed as a risky prospect. Individuals exhibiting inconsistent preferences across frames, and at odds with expected marginal utility are susceptible to the framing bias [22]. Furthermore, the choice of our material in this study is as a result of its robustness to alterations in contextual environments [23]. Subsequently, we collected demographic data for control purposes.

The independent variable is the framing bias while the dependent variable is the auditors' accurate detection of aggressive financial reporting items. HPSS, and prior experience in related experiments are control variables. The eye-tracking metric is the duration of fixations. In this study, the total duration of fixations is the total length of time of fixations of a participant on all the six AOIs, and it is a proxy for cognitive load, and processing levels [24]

Apparatus and Measures

Eye-tracking (Tobii pro nano) was used to gather the behavioral measures throughout the experiment, at a sampling frequency of 60 Hz. The total duration of fixations were gathered for each area of interest (AOI). Due to the randomization on the page of account examination, AOIs could have different representations per participant and per attempt. AOIs were all of the same size. For each participant, the eye tracker was calibrated using a nine-point fixation technique thus adjusting for participants' individual differences in eye characteristics [25].

Results

We briefly present some results from our study. Hypothesis 1 states that auditors are susceptible to the framing bias. As earlier indicated, individuals exhibiting inconsistent preferences across frames (neither consistently risk lovers nor consistently risk averse in the decision points), and whose preferences are at odds with expected marginal utility are susceptible to the framing bias. It was observed that 57.5% of participants were susceptible to the framing bias, compared to 42.5% who were not susceptible. This supports the assertion that auditors are susceptible to the framing bias.

Hypothesis 2 stipulates that the framing bias is associated with a higher total duration of fixations. A major objective of this study is to identify the differences in the attentional characteristics associated with the framing bias in a simple aggressive reporting detection task. To ascertain this, we verify if there is a significant difference between the total duration of fixations of biased participants and unbiased participants. We conduct a linear regression with random intercept model and a two-tailed level of significance. For all stimuli, we find that total duration of fixations were higher for biased participants than for unbiased participants (t=2.07, *p*-value=0.05). Again, the average duration of fixations were higher for the biased participants than the unbiased participants (t=2.369, *p*-value=0.023). Regarding only target stimuli (aggressive financial reporting items), we find that the total duration of fixations was higher for biased participants than for unbiased participants at a 10% significance threshold (t=1.90, *p*-value=0.07). Following this, we verify if the framing bias leads to poorer decision making.

Hypothesis 3 states that the framing bias leads to a poorer detection of aggressive financial reporting items. In testing H3, we control for the participant's trait skepticism (HPSS), and prior experience of related tests. Results of the model indicate an estimate of -1.35, and a *p*-value of 0.03 indicating that the framing bias is associated with poorer detection of aggressive financial reporting. We carry out a robustness check for multicollinearity and obtain variance inflation factors of 1.31 and 1.29 for the experiment experience, and the HPSS respectively, confirming the absence of multicollinearity among the independent variables.

Discussion and Conclusion

Our results suggest that H1 and H3 are fully are supported while H2 is fully supported for all stimuli and marginally supported for target stimuli. First, novice auditors are susceptible to the framing bias. With regards to the visual behavior of the biased auditors in the audit task, they are associated with higher total duration of fixations for all stimuli, and for target stimuli. The total duration of fixations is a proxy for processing levels. Finally, we find that the framing bias is associated with poorer detection of aggressive financial reporting elements.

Dual process theory (DPT) [26] provides a theoretical backing to our findings. The literature generally identifies three perspectives to DPT with regards to the order in which Type 1 and Type 2 processing occur [27]; default interventionist, parallel processing, and hybrid model. Our results support a hybrid model which proposes that Type 1 processing provides two potential intuitive responses which may then be moderated by Type 2 processing [28]. In our study, the more rapid and intuitive response which could be interpreted as Type 1, demonstrated by a lower total duration of fixations, generally led to correct responses. For biased individuals, this decision making process was moderated by a slow, reflective, higher processing effort which could be interpreted as Type 2, demonstrated by a higher total duration of fixations. In this case the apparent higher processing effort led to less accurate decisions since it was informed by a cognitive bias.

Our research so far offers both theoretical and practical contributions. On the theoretical side, our results support the idea that poorer detection of aggressive items is due to the interference of higher processing levels, possibly representing cognitive load or confusion as a result of the framing bias. However, we find no pupillometric evidence. While an alternative interpretation could be that the framing bias leads to more interest or focus on target information due to higher processing levels, the eventual outcome of a poorer detection of target stimuli invalidates this explanation. From a practical standpoint, this paper underscores the need, in the pursuit of high quality audits, to understand the susceptibility profiles of novice auditors to cognitive biases. Audit managers and seniors should collaborate with UX developers to elaborate personalized AMIS interfaces which present a lesser use of cognitive resources to novices who are more susceptible to cognitive biases.

Our study presents some limitations which provide the opportunity for further research. First, the homogeneous nature of the sample may not take cultural differences, a factor which may influence the effects of cognitive biases [29], into account. Consequently, individuals practicing in different jurisdictions may experience different outcomes. Further research could investigate these effects on non-European settings. Second, our study focuses on trait framing bias replicated from an original study in psychology. Further studies could adopt other manifestations of the framing bias to verify the generalizability of our conclusions.

Acknowledgement. This research was funded in whole or in part by the National Research Agency (ANR), France, under the project ANR-21-CE26-0012-01

Appendix: Experimental Material

Informed Consent form :
Dear Participant,
This study was developed as part of a research program conducted by X University in collaboration with researchers at Y University. It deals with practices related to financial auditing.
Your answers will remain strictly anonymous and will only be used for academic purposes. The accuracy and sincerity of your answers are crucial to the quality of this work. We thank you in advance for your kind cooperation.
Informed consent form This study attempts to gather information on the differences in individual performances during auditing tasks among professionals. You will be presented with a series of questions about an auditing task, your preferences, and your personality. The questionnaire lasts about 15 minutes. The risks of participation are minimal in this study. However, you may feel emotionally uncomfortable when you have to make judgments. We hope that thanks to your participation, researchers at X University and Y University will know more about the relationship between contextual and personal factors impacting performance during auditing tasks. All data obtained from participants will be kept confidential and will only be reported in a global format

(ie only combined results and never individual reports on a particular person). All the questionnaires will be anonymous and know that the research team will have access to them. The collected data will be stored on a secure server of the Qualtrics company until the principal investigator removes them. There is compensation for complete and valid participation. You should have validated all attention checks to receive compensation. Participation in this study is entirely voluntary. You have the right to withdraw at any time or refuse to participate fully. If you wish to withdraw, please inform the principal researcher at this email address: xxx. If you have any questions about this study, you can contact the principal researcher. x University's Ethics Board has determined that the data collection related to this study meets the ethics standards for research involving humans. If you have any questions related to ethics, please contact the Research and Ethics Board secretariat at xxx or by e-mail at xxx
I consent to participate in this study a.Yes b.No

Audit Task

You will now proceed to a self-paced review of audit evidence of Meter-Tek Company reported in 6 sentences, categorized into one of various financial statement accounts.
Meter-Tek is a manufacturer and marketer of water, electricity and natural gas meters and you are their auditor. Materiality as with other audits is set at $100,000.
Meter-Tek's accounting year is from 1st January to 31st December. The accounting year being audited is 2021.
The audit evidence will be displayed one at a time
Cash: The staff accountant noted that bank accounts are reconciled monthly
Trade Receivables: An examination of year-end customer balances indicates that the December 31, 2021 allowance for doubtful accounts is inadequate.
R&D and Engineering Expenses: Total engineering expenses decreased by $40,000 from 2020
Inventories: Test counts conducted at the December 31, 2021 inventory observation did not reveal exceptions and were subsequently agreed to the final inventory listing.
Investments in Affiliated Companies: Meter-Tek continues to hold equity interests of 25% in two profitable companies that are accounted for using the equity method.
Accounts Payable and Accrued Liabilities: The search for unrecorded liabilities involved an examination of payments and invoices processed subsequent to year-end and revealed significant understatements.

Audit Task Questions

Please evaluate the client's financial reporting as a whole.
Aggressive financial reporting refers to accounting practices that are designed to overstate a company's financial performance. It includes but is not limited to

1. Sharp rises in incomes or sharp decreases in expenses from previous years
2. Manipulations or violations of accounting principles, policies or standards to enhance financial performance
3. Misreporting

Not aggressive Very at all 1 2 3 4 5 6 7 8 9 10 aggressive

Of the following 6 accounts you have read on the previous page, which warrant further examination?

a. Cash
b. Accounts Payable and Accrued Liabilities
c. Trade Receivables
d. R&D and Engineering Expenses
e. Inventories
f. None
g. Investments in Affiliate Companies

Hurtt's Professional Skepticism Scale

Statements that people use to describe themselves are given below. Please circle the response that indicates how you generally feel. There are no right or wrong answers. Do not spend too much time on any one statement.

	Strongly Disagree					**Strongly Agree**
I often accept other people's explanations without further thought	1	2	3	4	5	6
I feel good about myself	1	2	3	4	5	6
I wait to decide on issues until I can get more information	1	2	3	4	5	6
The prospect of learning excites me	1	2	3	4	5	6
I am interested in what causes people to behave the way that they do.	1	2	3	4	5	6
I am confident of my abilities.	1	2	3	4	5	6
I often reject statements unless I have proof that they are true	1	2	3	4	5	6
Discovering new information is fun	1	2	3	4	5	6
I take my time when making decisions	1	2	3	4	5	6
I tend to immediately accept what other people tell me.	1	2	3	4	5	6
Other people's behavior does not interest me	1	2	3	4	5	6

I am self-assured.	1	2	3	4	5	6
My friends tell me that I usually question things that I see or hear	1	2	3	4	5	6
I like to understand the reason for other people's behavior	1	2	3	4	5	6
I think that learning is exciting.	1	2	3	4	5	6
I usually accept things I see read or hear at face value	1	2	3	4	5	6
I do not feel sure of myself	1	2	3	4	5	6
I usually notice inconsistencies in explanations	1	2	3	4	5	6
Most often I agree with the others in my group	1	2	3	4	5	6
I dislike having to make decisions quickly	1	2	3	4	5	6
I have confidence in myself	1	2	3	4	5	6
I do not like to decide until I've looked at all of the readily available information	1	2	3	4	5	6
I like searching for knowledge	1	2	3	4	5	6
I frequently question things that I see or hear	1	2	3	4	5	6
It is easy for other people to convince me	1	2	3	4	5	6
I seldom consider why people behave in a certain way	1	2	3	4	5	6
I like to ensure that I've considered most available information before making a decision	1	2	3	4	5	6

I enjoy trying to determine if what	1	2	3	4	5	6
I read or hear is true	1	2	3	4	5	6
I relish learning	1	2	3	4	5	6
The actions people take and the reasons for those actions are fascinating	1	2	3	4	5	6

Framing Bias

Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

Decision (i). Choose between: A. a sure gain of $240 B. 25% chance to gain $1000, and 75% chance to gain nothing

Decision (ii). Choose between: C. a sure loss of $750 D. 75% chance to lose $1000, and 25%

Demographic Questions

1. What is your gender? a. Male b. Female

2.In which age range (in years) are you? 18-20; 21-25; 26-30; 31-35; 36-40; 41-45; 46-50; 51-55; 56-60; 61-65; 66-70; 71-75; 76-80; 81-85

3. What is the highest level of education you have attained? a. No higher education degree b.Undergraduate c.Graduate d.PhD

4. What is your undergraduate major? Finance; Economics; Accounting; Marketing; HRM; Strategy; Supply Chain/logistics; Management; Other

5. Do you have any audit work experience (including internships)? Yes; No

6. Do you currently any accounting or auditing professional designation? CA; CGA;CMA; CPA; CFA; No

7. Prior to this experiment, have you participated in either accounting, finance, auditing, economics, or psychology experiments? Yes; No

References

1. Gajewski, JF., Heimann, M., Léger, PM., Teye, P. Nudging to Improve Financial Auditors' Behavior: Preliminary Results of an Experimental Study. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, PM., Randolph, A.B., Fischer, T. (eds) Information Systems and Neuroscience. NeuroIS 2020. Lecture Notes in Information Systems and Organisation, vol 43. Springer, Cham. (2020)
2. Erfanian, M., L.Meunier, and J.-F.Gajewski. "The Effect of Cognitive Load and Default Nudges on Decision-Making: Experimental Evidence From a Behavioral Study." Journal of Behavioral Decision Making39, no. 2: e70081. (2026)
3. Munidewi, I. A. B., Ludigdo, U., Djamhuri, A., and Andayani, W. Role of affective neuroscience in audit judgement and decisions making: A systematic

literature review for auditing research. Australasian Accounting Business & Finance Journal, 18(1), 130-147. (2024)

4. Mock, T. J. and H. Fukukawa (2016). Auditors' risk assessments: The effects of elicitation approach and assertion framing. Behavioral Research in Accounting 28(2), 75 – 84.
5. Kahneman, D. and A. Tversky. Choices, values, and frames. American Psychologist 39(4), 341–350. (1984)
6. Fukukawa, H. and T. J. Mock (2011). Audit risk assessments using belief versus probability. Auditing: A Journal of Practice & Theory 30(1), 75 – 99.
7. Griffith, E. E., Kadous, K., & Young, D. Improving Complex Audit Judgments: A Framework and Evidence. Contemporary Accounting Research, 38(3), 2071–2104 (2021)
8. Lynch, E. J. and L. M. Andiola. If eyes are the window to our soul, what role does eye-tracking play in accounting research?. Journal of Information Systems 33(3), 107 – 133. (2019)
9. Gilovich, T. and D. Griffin. Introduction – heuristics and biases: then and now", in Gilovich, T., Griffin, D. and Kahneman, D. (Eds), Heuristics and Biases. Cambridge University Press, New York, pp. 1-18. (2013)
10. Shanteau, J. Cognitive heuristics and biases in behavioral auditing: Review, comments and observations. Accounting, Organizations & Society 14(1/2), 165 – 177. (1989)
11. Knapp, M. C. and C. A. Knapp. Cognitive biases in audit engagements. CPA Journal 82(6), 40 – 45 (2012)
12. Henrizi, P., D. Himmelsbach, and S. Hunziker. Anchoring and adjustment effects on audit judgments: experimental evidence from Switzerland. Journal of Applied Accounting Research 22(4), 598 – 621 (2021)
13. McMillan JJ, White RA.: Auditors' Belief Revisions and Evidence Search: The Effect of Hypothesis Frame, Confirmation Bias, and Professional Skepticism. Accounting Review 68(3):pp. 443-465 (1993)
14. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye Tracking: A Comprehensive Guide to Methods and Measures. OUP, Oxford (2011)
15. Goldberg, J.H., Kotval, X.P.: Computer interface evaluation using eye movements: methods and constructs. Int. J. Ind. Ergon. **24**(6), 631–645 (1999)
16. Asare, S. K. The auditor's going-concern decision: Interaction of task variables and the sequential processing of evidence. The Accounting Review 67(2), 379-393. (1992)
17. Pearson, C. A. The effects of work division and technology on auditor engagement and performance. International Journal of Auditing, 29(1), 229–241. (2025)
18. Phillips, F.: Auditor Attention to and Judgments of Aggressive Financial Reporting. Journal of Accounting Research,37(1), 167-189. (1999).
19. Gajewski, J. F., Heimann, M., Léger, P. M., & Teye, P.. Enhancing auditors' professional skepticism through nudges: an eye-tracking experiment. Accounting and Business Research, 55(4), 426–444. (2025)

20. Brosnan, Kylie, Bettina Grün, and Sara Dolnicar. Cognitive Load Reduction Strategies in Questionnaire Design. International Journal of Market Research 63 (2): 125–33. (2021)

21. Hurtt RK.: Development of a Scale to Measure Professional Skepticism. Auditing 2010 29(1) pp. 149-171. (2010)

22. Tversky, A. and D. Kahneman. The framing of decisions and the psychology of choice. Science 211(4481), 453–458. (1981)

23. Wolaver, Amy M., and Doces, John A... "The Impact of COVID-19 and Political Identification on Framing Bias in an Infectious Disease Experiment: The Frame Reigns Supreme." Social Science Quarterly 102 (6): 2459–71. (2021)

24. Yusuf, S., Kagdi, H., Maletic, J.I.: Assessing the comprehension of UML class diagrams via eye tracking. In: Program Comprehension, 2007. ICPC'07, 15th IEEE International Conference, pp. 113–122 (2007)

25. Rose, A. M., J. M. Rose, K. Rotaru, K.-A. Sanderson, and J. C. Thibodeau. Effects of uncertainty visualization on attention, arousal, and judgment. Behavioral Research in Accounting 34(1), 113 – 139. (2022)

26. Kahneman, Daniel. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux. (2011)

27. Lawrie, E., Flus, M., Olechowski, A., Hay, L., & Wodehouse, A. From theory to practice: a roadmap for applying dual-process theory in design cognition research. *Journal of Engineering Design*, *36*(12), 2204–2224. (2025)

28. Neys, Wim De. "Bias, Conflict, and Fast Logic." In *Dual Process Theory 2.0*, edited by Wim De Neys.47–65. New York: Routledge. 2018

29. Loibl, C., C. R. Sunstein, J. Rauber, and L. A. Reisch. Which europeans like nudges? approval and controversy in four european countries. The Journal of Consumer Affairs 52(3), 655–688. (2018)

From Experience to Biomarkers: Surveying Neurophysiological Measures and Correlates of Flow

Fabian J. Stangl[1] and Thomas Kalischko[2]

[1] University of Applied Sciences Upper Austria, Steyr, Austria
fabian.stangl@fh-steyr.at
[2] University of Applied Sciences Wiener Neustadt, Wieselburg, Austria
thomas.kalischko@fhwn.ac.at

Abstract. Flow is a central concept for understanding optimal, intrinsically rewarding experiences during task engagement. This paper reviews the use of Neuro-Information Systems (NeuroIS) measurements to study flow and its neurophysiological correlates. Based on a literature review of 63 papers, four measurement domains were identified: autonomic nervous system (ANS) measures, brain imaging methods, hormone-related measurements, and multimodal approaches. Brain imaging methods, particularly electroencephalography (EEG), were the most frequently used approach, followed by ANS-related measures such as heart rate (HR) and heart rate variability (HRV). Hormone-related approaches were rare, whereas multimodal designs show promise by integrating complementary physiological perspectives and distinguishing flow from related states. Across domains, neurophysiological measurements were mainly used to examine correlates of self-reported flow, validate induced flow states, or predict flow-related states rather than to measure flow directly. This review provides a valuable foundation for future NeuroIS research on flow.

Keywords: Flow · Flow State · Neuro-Information Systems (NeuroIS) · Neurophysiological Measures · Literature Review

Introduction

Flow has become a central concept for explaining optimal, intrinsically rewarding experiences during task engagement [1]. In information systems research, flow and related constructs, such as cognitive absorption [2], capture episodes of deeply engaged technology use [3]. This state is of particular interest because it is associated with sustained engagement and intrinsic motivation [e.g., 4]. It has also been linked to consequential outcomes in digital work settings, such as experiential outcomes during technology use (e.g., immersion; [5]) or task performance [e.g., 6].

Flow is typically described as a state of optimal functioning that is temporally bounded, in which attention is strongly focused on the ongoing activity, the experience is inherently enjoyable, and individuals perceive a pronounced sense of control over their actions [1, 7, 8]. Theoretically, flow is most likely to emerge when perceived task challenges are well-matched to perceived skills, resulting in sustained involvement and the seamless coordination of action and awareness [1, 3, 9]. However, these defining characteristics also make flow difficult to study empirically. Due to its development and fluctuation during task execution, as well as its inaccessibility to conscious reflection in real time, post-hoc self-reports often provide only a coarse summary, blurring the onset, duration, and intensity of a flow episode [10]. Therefore, this reliance on retrospective measurement limits temporal precision and can introduce recall- and evaluation-related bias, particularly in realistic work settings where flow may be intermittent or repeatedly disrupted [e.g., 6].

Neuro-Information Systems (NeuroIS) provides an interdisciplinary approach to studying interactions with digital technologies and the underlying cognitive, emotional, and behavioral processes involved [11–14]. Methodologically, NeuroIS draws on a broad range of neurophysiological tools, including indicators of the autonomic nervous system (e.g., heart-related measures), brain imaging methods such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), and, in some cases, neuroendocrine markers [14, 15]. These methods are particularly promising for flow research because flow is dynamic and partly pre-reflective [10]. Neurophysiological signals can provide time-resolved indicators of processes associated with task engagement, thereby complementing post hoc self-reports [10, 14–16].

Despite the growing body of literature on flow and its neurophysiological correlates, the evidence base remains fragmented. Existing work ranges from methodological contributions reviewing practices in neurophysiological flow studies [e.g., 10] to more selective perspectives focusing on specific signal families (e.g., peripheral measures; [17]). A recent review highlights the broader potential of physiological measurements in flow research, emphasizing their capacity for continuous, non-intrusive assessment of flow dynamics during activities [16]. However, these reviews do not provide an overview of how NeuroIS measurements have been used in different neurophysiological domains to study flow and its correlates. Prior work either focuses on a specific class of measures [e.g., 17] or addresses flow research more broadly [e.g., 16], rather than examining NeuroIS measurement approaches in a targeted way. In addition, related work has explored the thematic structure of this literature [18]. As a result, it is difficult to compare studies and gain an overall picture of the neurophysiological measures used in the flow literature. Against this backdrop, this paper aims to provide an overview of NeuroIS measures in flow research, addressing the following research question: **How has existing research used NeuroIS measurements to examine flow and its neurophysiological correlates?** To answer it, we provide a methodological overview of the neurophysiological measures used in prior studies and discuss implications for future NeuroIS flow research.

Review Methodology

To examine the scope, range, and nature of existing research applying NeuroIS measurements to investigate flow and its neurophysiological correlates, we conducted a scoping review of the literature [19–21]. The review followed established methodological guidelines for literature searches [22, 23] and included peer-reviewed journal articles and conference proceedings published in English with no restrictions on the year of publication.

Our literature review process consisted of three phases: identification, screening, and selection. Overall, we identified a total of 63 relevant papers, which formed the basis for our examination of how NeuroIS measurements have been applied to the study of flow and its neurophysiological correlates.

Phase 1: Literature Identification – The starting point for our review was the systematic literature review by Knierim et al. [17], which synthesized prior research on psychophysiological flow measurement with a particular focus on indicators of the peripheral nervous system. This work provided an important conceptual and methodological foundation for our review, especially in identifying relevant terminology related to neurophysiological flow measurement. To capture the broader NeuroIS literature, we also drew on the work of Riedl et al. [24], who examined the development of the NeuroIS research field and outlined relevant NeuroIS terminology. Together, these studies informed our keyword selection and ensured that our review adhered to the established principles of conducting systematic literature reviews [22, 23].

For our literature search, we used search terms that represented the two core dimensions of this review: flow and neurophysiological measurements related to NeuroIS. The flow-related search string captured the focal phenomenon with terms such as "flow state", "psychological flow", "flow experience", and "state of flow". To capture the neurophysiological dimension, we used a broad set of NeuroIS-related terms covering general concepts, such as "Neuro-Information Systems", "NeuroIS", "neurophysiology", "psychophysiology", and "physiology" as well as specific measurement domains and signal types. These included the autonomic and central nervous systems, related recording methods, biomarkers, and other neuroendocrine indicators. To ensure comprehensive coverage, we incorporated synonyms, abbreviations, and wildcard operators, which allowed us to identify a broad set of potentially relevant papers.[4]

To address the phenomenon of interest (i.e., flow) and the measurement perspective (i.e., neurophysiological measures related to NeuroIS), we adopted the research methodology of Knierim et al. [17] and conducted a literature review using two primary academic databases: Scopus and Web of Science. The database search was conducted on March 3, 2026. This approach ensured broad coverage of peer-reviewed journal articles and conference proceedings relevant to our review. Moreover, we manually searched all NeuroIS Retreat proceedings [25–35] to identify additional relevant papers on flow and its neurophysiological measurement or correlates. Through this process, we identified 2,183 papers, forming the basis for subsequent screening and selection.

---

[4] The final search string was as follows: *("flow state" OR "psychological flow" OR "flow experience" OR "state of flow") AND ("Neuro-Information Systems" OR NeuroIS OR neurophysiolog* OR psychophysiolog* OR physiolog* OR "autonomic nervous system" OR ANS OR "central nervous system" OR CNS OR EEG OR electroencephalograph* OR fMRI OR "functional magnetic resonance imaging" OR MEG OR magnetoencephalograph* OR NIRS OR fNIRS OR "functional near infrared spectroscopy" OR ECG OR electrocardiogra* OR EMG OR electromyogra* OR EDA OR "electrodermal activity" OR "skin conductance" OR galvan* OR HRV OR "heart rate variability" OR "heart rate" OR "eye tracking" OR pupillometr* OR oculometr* OR hormone* OR cortisol OR neuroendocrine OR saliva* OR urine* OR blood*).*

Phase 2: Literature Screening – During this phase, we applied the inclusion criteria to the titles, abstracts, and keywords of the identified papers. To do so, we adapted the study selection criteria proposed by Knierim et al. [17]. To be included in this review, a study had to meet three criteria. First, the study had to be published as a peer-reviewed journal article, article in press, conference paper, or book chapter. Second, the study had to contain an empirical component involving a neurophysiological measurement to investigate flow or its correlates. Third, the study had to be relevant to the NeuroIS domain. Applying these criteria at the title, abstract, and keyword levels resulted in the exclusion of 2,105 papers.

Phase 3: Literature Selection – During this phase, we evaluated the full texts of the remaining 78 papers using the same inclusion criteria. Of these, 12 were removed as duplicate records identified through multiple data sources. We then excluded an additional 20 papers that did not meet the inclusion criteria. Specifically, we excluded non-peer-reviewed papers [e.g., 36], non-empirical papers (e.g., methodological [e.g., 37] and conceptual papers [e.g., 38]), and studies that investigated flow using self-report measures only [e.g., 39–42] or that were irrelevant to our research objective [e.g., 43, 44] or to NeuroIS [e.g., 45, 46]. This process resulted in 46 relevant papers. To further expand the literature base, we followed the established guidelines for literature reviews [22, 23] and subsequently conducted backward and forward searches based on the retained papers. Through this process, we identified an additional 20 relevant papers [e.g., 6, 47]. As a final step, we merged papers that referred to the same empirical study. Specifically, we merged the studies by Bian et al. [48, 49], Drachen et al. [50, 51], and Nacke and Lindley [52, 53], retaining the more developed version in each case [i.e., 48, 51, 53]. As a result, our final literature base included 63 papers on flow and its correlates.

Compared with previous literature reviews, our methodology provides broader and more comprehensive coverage of the neurophysiological measures and correlates used to study flow. For comparison, recent reviews analyzed 20 studies that focused on the peripheral nervous system [17] and 20 studies that focused on physiological and/or neuropsychological methods for studying flow [16].

Review Results

This section presents the main results of the literature review. To provide a structured overview of the identified studies, we classified the papers according to their measurement domain and method. In terms of measurement domain, we distinguished between studies using indicators of the autonomic nervous system (ANS), brain imaging methods, and hormone-related measurements. Moreover, we differentiated between studies that relied on a single measurement method, multiple methods within the same domain, and multiple methods across different domains. This classification approach is based on the categorization of NeuroIS tools proposed by Riedl and Léger [15; pp. 47-72] and is consistent with its application in recent reviews in interruption science [54–57]. Taken together, this framework provides an overview of how neurophysiological measures and their correlates have been applied in flow research.

Overall, our analysis (see **Table 1**) shows that brain imaging methods were the most frequently used approach to study flow and its correlates (46%), followed by ANS measures (33.3%), multimodal approaches (17.5%), and hormone-related measurements (3.2%). Below, we synthesize the findings within each measurement approach.

Table 1. Overview of NeuroIS Measurements and Correlates in Flow Research

Domain	Measurement Method	Papers	Reference(s)
Autonomic Nervous System (ANS)		**21**	
	HR and HRV	6	[58–63]
	HRV	4	[6, 64–66]
	EDA & EMG	2	[53, 67]
	EDA & HR	2	[51, 68]
	Automated Facial Expression Analysis, EDA, HR & Respiration	1	[69]
	EDA, HR & Respiration	1	[70]
	EDA, HR and HRV & Respiration	1	[71]
	EMG, HR and HRV & Respiration	1	[48]
	Eye Tracking	1	[72]
	Eye Tracking, HR and HRV	1	[73]
	HR and HRV, Facial Video, Posture & Speech	1	[74]
Brain Imaging Methods		**29**	
	EEG	21	[75–95]
	fMRI	3	[96–98]
	fNIRS	3	[99–101]
	EEG & fNIRS	1	[102]
	PET	1	[103]
Hormone-Related Measurements		**2**	
	Saliva	2	[104, 105]
Multimodal Approaches		**11**	
ANS & Brain Imaging		**9**	
	HR and HRV, Respiration & fNIRS	2	[106, 107]
	Blood Oxygen Saturation (SpO_2), EDA, HR and HRV & EEG	1	[108]
	Blood Volume Pulse, EDA, EMG, Eye Tracking, HR, Respiration & EEG	1	[109]
	EDA, HR & EEG	1	[110]
	EDA, HR and HRV & EEG	1	[47]
	EDA & fMRI	1	[111]
	HR & EEG	1	[112]
	HR and HRV, Respiration & EEG	1	[113]
ANS & Hormone-Related Measurements		**2**	
	HRV & Saliva	2	[114, 115]

Measurement Approach 1: Autonomic Nervous System – ANS measures were the second most frequently used measurement domain in the reviewed literature. In total, 21 studies used ANS-based approaches to examine flow and its correlates. Most of these studies relied on heart-related measures, including studies combining heart rate

(HR) and heart rate variability (HRV) (n = 6) [58–63], as well as studies focusing on HRV only (n = 4) [6, 64–66]. An additional study used HR and HRV alongside facial videos, upper body posture, and speech data in a video learning context [74]. Other studies used combinations of electrodermal activity (EDA) and electromyography (EMG) (n = 2) [53, 67], EDA and HR (n = 2) [51, 68], automated facial expression analysis, EDA, HR, and respiration [69], EDA, HR, and respiration [70], EDA, HR, HRV, and respiration [71], EMG, HR, HRV, and respiration [48], eye tracking only [72], or eye tracking together with HR and HRV [73]. Overall, these findings suggest that research on ANS-based flow is dominated by cardiovascular measures. Electrodermal, respiratory, facial, muscular, and ocular indicators have been used less frequently, primarily in conjunction with cardiac signals.

Across the HR- and HRV-based studies, a recurring pattern emerged. These measures were primarily used to examine the physiological correlates of self-reported or experimentally induced flow [6, 58–65], to compare flow with adjacent states such as boredom and anxiety [63] or with flow under interruptions [6], and, in some cases, to predict flow-related states from physiological data [62, 65, 74]. Moreover, several studies highlight the importance of different heart rate variability indicators that reflect sympathetic and parasympathetic regulation when examining autonomic balance during flow-related states [6, 58–65]. Other ANS-based studies extend this perspective beyond cardiac measures alone. EDA- and EMG-based studies link flow to high-arousal, positive gameplay experiences [53, 67]. Furthermore, research combining EDA, heart rate, and respiratory signals indicates that ANS profiles incorporating multiple markers can distinguish flow from boredom, anxiety, and frustration, thereby supporting classification approaches in gaming contexts [48, 51, 68–71]. Similarly, eye-tracking studies indicate that ocular measures, either alone [72] or in combination with HR and HRV [73], can contribute to assessing attentional and effort-related aspects of flow. In addition, Wang et al. [74] used multimodal data, including HR, HRV, facial videos, posture, and speech, to classify states of boredom, flow, and anxiety. They also used this data to predict self-reported flow experiences during video learning. Together, these studies demonstrate that ANS research has primarily considered physiological signals as correlates, comparators, or predictors of flow-related states rather than as independent operationalizations of flow itself.

At the same time, the reviewed studies differ in their interpretation of the exact autonomic signature of flow. Some studies associate flow with increased sympathetic activation [48, 58, 71], reduced HRV [63], or moderate physiological strain [58, 71], whereas others interpret flow in terms of parasympathetic balance [59, 73], efficient regulation [64], or mixed patterns that depend on task and context [6, 60]. Comparability is further limited by substantial methodological variation. The studies differ in context, ranging from daily activities [59] and driving [63, 73] to office [6] and knowledge work [64], human–robot collaboration [66], and video [51, 53, 62, 67–69, 71] and virtual reality games [48, 60, 61], and video learning [74]. The studies also differ in design. Some rely on tightly controlled laboratory manipulations [6, 58, 63, 71, 73], some use more naturalistic or in-field measurements [59, 65], and some apply machine-learning approaches to classify flow-related states [62, 68, 72, 74]. Moreover, not all studies examine flow in the same way. Some focus directly on self-reported flow [6, 48, 58–

65, 69–73], while others emphasize related constructs or antecedents, such as executive task performance [58], time perception [61], affect recognition [62], interruption relevance [6], perceived challenge [66], or immersion and gameplay experience [51, 53, 67, 68].

Taken together, ANS-based measures offer the clearest insight into the peripheral physiological correlates of flow, particularly cardiovascular and arousal-related dynamics. These measures appear well suited to supporting ecologically valid [116, 117] assessments of flow-related states across work, learning, and gaming contexts.

Measurement Approach 2: Brain Imaging Methods – Brain imaging methods were the most frequently represented measurement domain in the reviewed literature. In total, 29 studies used brain imaging approaches to examine flow and its correlates. Most of these studies relied on EEG (n = 21) [75–95], followed by fMRI (n = 3) [96–98] and functional near-infrared spectroscopy (fNIRS) (n = 3) [99–101]. Moreover, one study combined EEG and fNIRS [102], while another used positron emission tomography (PET) to examine the neurochemical correlates of flow proneness [103]. Overall, these findings suggest that EEG dominates brain-imaging-based flow research, while fMRI, fNIRS, PET, and combinations of EEG and fNIRS are used less frequently for more specific analytical purposes.

Across EEG studies, a recurring pattern emerged. Brain imaging was primarily used to identify the neural correlates of self-reported or experimentally induced flow [77, 85, 88, 90, 94], to distinguish flow from related states such as boredom or overload [81, 83, 90], and to predict or monitor flow intensity based on neural signals [76, 77, 83, 84]. Moreover, several papers point to increased frontal theta activity [77, 90, 91], moderate or task-dependent alpha activity [77, 90, 92], and differentiated beta-band activity [77, 90] as promising neural correlates of flow-related states. A substantial subset of EEG studies has also explored prediction and classification, including real-time assessment [76, 83], dynamic monitoring [84], and flow-state detection in learning contexts [88]. Other EEG studies have demonstrated the feasibility of low-cost [85], single-channel [89], wearable [76], around-the-ear implementations [91]. Emerging research further points to artificial intelligence-driven personalization and flow-related neurophysiological assessment in learning contexts [82]. Adjacent EEG instrumentation work has explored 3D-printed dry-electrode systems for cognitive load monitoring in knowledge work [93]. This work may inform future flow research, but it does not constitute direct evidence of flow itself. The broader picture beyond EEG is complementary. For example, fMRI studies have linked flow to increased activation in regions associated with task engagement and reward, such as the inferior frontal gyrus, putamen, and midbrain [96–98]. These studies have also linked flow to decreased activation in regions linked to self-referential and emotional processing, including the medial prefrontal cortex and amygdala [97, 98]. Similarly, fNIRS studies have highlighted the importance of prefrontal activity in individual flow experience [100, 101] and in shared flow dynamics during collaborative learning [99]. Furthermore, a combined EEG-fNIRS study indicates that predicting flow intensity with multiple markers is possible [102], while a PET study broadens the perspective by linking dispositional flow proneness to dopamine D2-receptor availability in the dorsal striatum [103].

At the same time, the reviewed studies disagree about the exact neural signature of flow and the conditions under which it emerges. Even within the EEG literature, the reported correlates are inconsistent. Some studies emphasize increased frontal theta and moderate alpha patterns during flow experiences [77, 90], while others highlight delta and gamma activity [89], upper alpha stability [92], or quadratic relationships between flow and neural activity [91]. These discrepancies depend on the task and setting. Comparability is further limited by substantial methodological variation. The studies differ in context, ranging from video games [77, 79, 81, 83, 84, 96, 100, 101] and e-learning [85, 86, 88] to knowledge work [75, 91–93], website design [95], and collaborative learning [99]. The studies also differ in their measurement setups, ranging from low-cost [85], single-channel [89], wearable headband [76], around-the-ear EEG [91], and 3D-printed dry-electrode systems [93], to more controlled, multi-channel, laboratory recordings [77, 90, 94]. Moreover, studies examining flow do not all approach the topic in the same way. Some focus directly on self-reported flow [77, 85, 90, 94, 98, 100], while others emphasize related constructs or antecedents, such as engagement [79, 87], attention [85, 86, 88, 95], learning progress [78], optimal challenge or challenge–skill balance [80, 85, 86, 88], cognitive load [82, 93], or dispositional flow proneness [103].

Taken together, brain imaging methods offer the most direct insight into the neural and, in the case of PET, neurochemical correlates of flow. These methods also demonstrate great promise in predicting and evaluating flow-related states in real time [76, 77, 83, 84, 88, 102]. However, a single, stable, brain-based signature of flow has yet to emerge.

Measurement Approach 3: Hormone-Related Measurements – Hormone-related measurements were the least frequently represented domain in the reviewed literature. Only two studies examined flow in relation to cortisol [i.e., 104, 105], suggesting that endocrine approaches remain uncommon. However, these studies differed substantially in design and analytical focus. Brom et al. [104] investigated cortisol responses in a quasi-experimental serious-game setting and found that flow was positively related to positive affect and negatively related to negative affect, while showing almost no direct relationship between flow and cortisol levels. Elevated cortisol levels were primarily observed among socially anxious males in team-based conditions. In contrast, Peifer et al. [105] used a double-blind, randomized, placebo-controlled crossover design to experimentally manipulate cortisol and found that elevated cortisol reduced self-reported flow during a Pac-Man task.

Taken together, these studies suggest that cortisol has been used to examine endocrine correlates and possible causal effects on flow rather than to operationalize flow itself. Overall, the evidence base is small and methodologically heterogeneous, indicating that endocrine perspectives on flow remain underdeveloped.

Measurement Approach 4: Multimodal Approaches – Multimodal approaches constituted a smaller yet still notable share of the reviewed literature. In our review, 11 studies combined measures from more than one physiological domain. Most of these studies integrated ANS measures with brain imaging methods (n = 9) [47, 106–113]. A smaller subset combined ANS measures with hormone-related measurements, specifically HRV and salivary cortisol (n = 2) [114, 115]. Overall, these studies reflect an effort to capture flow and related states through complementary central and peripheral physiological perspectives rather than through a single signal source.

Across studies combining ANS and brain imaging methods, a recurring pattern emerged. Multimodal designs were primarily used to triangulate self-reported or experimentally induced flow and related states rather than to operationalize flow directly from physiology alone [47, 106–113]. Together, these studies imply that flow is linked to simultaneous alterations in cortical processing and peripheral physiology. These alterations include stronger activation of attention-related brain networks, specific EEG patterns, and simultaneous changes in HR/HRV, respiration, and EDA [106–108, 110–113]. Several studies also indicate that combining modalities can distinguish flow from adjacent states, such as boredom, overload, or frustration [108, 110, 111]. More recent work suggests that such multimodal assessments are increasingly feasible with wearable devices [108]. Two studies combining ANS and hormone-related measurements extend this perspective, indicating that flow is linked to a specific autonomic-endocrine profile. Preliminary evidence suggests that flow is associated with a balance between skills and demands and with the interplay of sympathetic, parasympathetic, and hypothalamic-pituitary-adrenal axis processes [114, 115].

At the same time, the reviewed studies differ in their interpretation of the exact physiological signature of flow. For example, fNIRS-based studies do not converge on a single cortical interpretation. De Sampaio Barros et al. [106] associate flow with increased oxygenated hemoglobin in the frontoparietal attention network, whereas Harmat et al. [107] report no association between flow and prefrontal oxygenation and find no evidence of hypofrontality. Peripheral findings are likewise mixed. Some studies point to reduced HRV or stress-related activation during flow-related states [114], whereas others emphasize deeper respiration, parasympathetic involvement, moderate rather than consistently elevated sympathetic and hypothalamic-pituitary-adrenal-axis activation, or no clear HR differences [107, 112, 115]. Comparability is further limited by substantial variation in context, task, and construct operationalization. The multimodal studies range from video games [108, 113] and Facebook use [109] to enterprise resource planning training [47] and robotic surgery [112], with some focusing on constructs related to flow, such as cognitive absorption [47] or a core flow state [109].

Taken together, multimodal approaches offer the most comprehensive view of the neurophysiological correlates of flow. This suggests that flow-related processes engage the autonomic and central nervous systems, and occasionally the endocrine system. Compared with single-modality studies, multimodal approaches have the advantage of being able to triangulate central and peripheral processes, distinguishing flow more robustly from related states, such as boredom, overload, and frustration. Nevertheless, the evidence remains heterogeneous and still relies heavily on self-report validation.

Discussion and Concluding Remarks

This review examined how existing research has used NeuroIS measurements to study flow and its neurophysiological correlates. The findings show that flow research draws on a broad range of NeuroIS approaches, including ANS-related measures, brain imaging methods, hormone-related measurements, and multimodal designs. At the same time, the review revealed a clear methodological pattern. Most studies do not operationalize flow exclusively through neurophysiological data. Instead, they use physiological signals to examine correlates of self-reported flow, validate experimentally induced flow conditions, or predict subjective, flow-related states. This pattern aligns with the conceptual nature of flow as a dynamic, temporally bounded, and partly pre-reflective experience that is difficult to access directly in real time [1, 7, 8, 10]. It also aligns with the broader NeuroIS perspective, which emphasizes the value of neurophysiological measures as a complement to, rather than a replacement for, established behavioral and self-report approaches [11–15, 118].

Beyond this general pattern, the review offers three additional, more specific contributions. First, it provides a structured overview of how the literature is distributed across NeuroIS measurement domains (see **Table 1**). Brain imaging methods were the most frequently used category, with EEG clearly dominating. ANS-related studies formed the second-largest group, relying primarily on cardiovascular indicators, particularly HR and HRV, which are often complemented by EDA, EMG, or respiration. Hormone-related approaches were rare, suggesting that endocrine perspectives remain underrepresented in NeuroIS flow research. Second, the review shows that despite substantial progress, the literature has not yet converged on a single, stable physiological signature of flow. Cross-study comparability is limited because the reviewed literature often investigates related constructs (e.g., cognitive absorption; [47]) or antecedents (e.g., dispositional flow proneness; [103]) rather than flow itself. Instead, findings vary depending on the context, task, measurement setup, and construct operationalization. Third, the review highlights multimodal approaches as particularly promising because they combine complementary physiological perspectives and are well-suited to the dynamic, multidimensional nature of flow. Thus, this review builds on previous work that examined broader developments in flow research [16], peripheral nervous system indicators [17], or the thematic structure of this research stream [18]. This review shows how different NeuroIS measurement domains have been used and where their findings converge or diverge. It also illustrates their collective implications for future flow research.

Taken together, these findings suggest substantial progress in identifying neurophysiological indicators of flow. However, research is far from directly operationalizing the construct physiologically in a consistent manner. This is not surprising given that flow is a holistic, experiential state characterized by absorption, enjoyment, control, and a balance between challenge and skill [1, 7, 8]. Such a state is unlikely to be adequately captured through a single signal or isolated physiological marker. The reviewed literature suggests that NeuroIS methods are valuable for improving temporal sensitivity, revealing underlying cognitive and affective processes during task engagement, and extending flow research beyond retrospective assessments [10–15, 118].

The review also yields several implications for future research. First, future studies should distinguish more clearly between attempts to directly operationalize flow through neurophysiological measurements and studies investigating physiological correlates of self-reported flow. Explicitly separating these approaches conceptually would improve comparability across studies and support the development of cumulative knowledge. Second, future ANS research would benefit from more consistent reporting and interpretation of cardiovascular indicators, particularly HR and HRV, because the reviewed studies vary substantially in how autonomic activation and regulation are conceptualized across tasks and contexts [58, 59, 63, 64]. Related reviews in NeuroIS likewise emphasize the need for greater clarity and consistency in the use of HR- and HRV-based measures [119–121]. Third, future brain-imaging research should more explicitly connect neural markers to specific flow dimensions, such as attention [85], control [78], and challenge-skill balance [88]. This research should also examine whether reported signatures replicate across comparable task settings. Fourth, hormone-related research [i.e., 104, 105] remains too limited to support firm conclusions, indicating the need for further studies before endocrine correlates of flow can be evaluated systematically. Fifth, multimodal approaches deserve particular attention. Compared with single-modality designs, they have greater potential to triangulate autonomic, central, and, in some cases, endocrine processes. They also distinguish flow more robustly from adjacent states [108, 110, 111], and benefit from recent advances in wearable sensing [108]. Finally, a temporal analysis of how these approaches have developed over time would strengthen this picture further, particularly in light of evolving sensing technologies and measurement opportunities. Although a full longitudinal analysis was beyond the scope of this paper, the reviewed literature suggests a recent increase in wearable [e.g., 108] and real-time measurement approaches [e.g., 83, 84], especially in EEG-based and multimodal studies. For example, recent research involving 3D-printed EEG hardware [93] illustrates this point. This development indicates growing interest in more ecologically valid [116, 117] and potentially adaptive [122] assessments of flow-related states.

This review should be interpreted in light of its limitations. As a scoping review [19–21], its purpose was to provide an overview of the literature, not to conduct a formal meta-analysis of effect sizes or measurement quality. Furthermore, the reviewed studies vary considerably in terms of context, task type, flow operationalization, and measurement setup, which limits the direct comparability of findings across studies. Despite these limitations, however, this review contributes to the NeuroIS literature by synthesizing how neurophysiological measurements have been used to study flow and its correlates, as well as by identifying dominant measurement approaches and promising directions for future research.

References

1. Csíkszentmihályi, M.: Beyond boredom and anxiety: Experiencing flow in work and play. Jossey-Bass Publishers, San Francisco (1975).
2. Agarwal, R., Karahanna, E.: Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly. 24, 665–694 (2000). https://doi.org/10.2307/3250951.
3. Pilke, E.M.: Flow experiences in information technology use. International Journal of Human-Computer Studies. 61, 347–357 (2004). https://doi.org/10.1016/j.ijhcs.2004.01.004.
4. Gerow, J.E., Ayyagari, R., Thatcher, J.B., Roth, P.L.: Can we have fun @ work? The role of intrinsic motivation for utilitarian systems. European Journal of Information Systems.

22, 360–380 (2013). https://doi.org/10.1057/ejis.2012.25.
5. Mütterlein, J., Berger, Benedikt Waltermann, H.: How head-mounted displays affect immersion in virtual reality experiences: A laboratory experiment. AIS Transactions on Human-Computer Interaction. 17, 343–368 (2025). https://doi.org/10.17705/1thci.00228.
6. Nadj, M., Rissler, R., Adam, M., Knierim, M., Li, M., Maedche, A., Riedl, R.: What disrupts flow in office work? The impact of frequency and relevance of IT-mediated interruptions. MIS Quarterly. 47, 1615–1646 (2023). https://doi.org/10.25300/MISQ/2023/17200.
7. Csíkszentmihályi, M.: Flow: The psychology of optimal experience. Harper Collins, New York (1990).
8. Csíkszentmihályi, M.: Finding flow: The psychology of engagement with everyday life. Basic Books, New York (1998).
9. Nakamura, J., Csíkszentmihályi, M.: The experience of flow: Theory and research. In: Snyder, C.R., Lopez, S.J., Edwards, L.M., and Marques, S.C. (eds.) The Oxford Handbook of Positive Psychology. pp. 279–296. Oxford University Press, New York (2021).
10. Durcan, O., Holland, P., Bhattacharya, J.: A framework for neurophysiological experiments on flow states. Communications Psychology. 2, 66 (2024). https://doi.org/10.1038/s44271-024-00115-3.
11. Dimoka, A., Pavlou, P.A., Davis, F.D.: NeuroIS: The potential of cognitive neuroscience for information systems research. Information Systems Research. 22, 687–702 (2011). https://doi.org/10.1287/isre.1100.0284.
12. 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.: On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS Quarterly. 36, 679–702 (2012). https://doi.org/10.2307/41703475.
13. 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.: On the foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. Communications of the Association for Information Systems. 27, 243–264 (2010). https://doi.org/10.17705/1CAIS.02715.
14. Riedl, R., Stangl, F.J.: NeuroIS. In: Baiyere, A., Galliers, R.D., and Stein, M.-K. (eds.) The Routledge Companion to Management Information Systems. pp. 198–216. Routledge, London (2025). https://doi.org/10.4324/9781032690483-18.
15. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS: Information systems and the brain. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-45091-8.
16. Peifer, C., Wolters, G., Harmat, L., Heutte, J., Tan, J., Freire, T., Tavares, D., Fonte, C., Andersen, F.O., van den Hout, J., Šimleša, M., Pola, L., Ceja, L., Triberti, S.: A scoping review of flow research. Frontiers in Psychology. 13, 815665 (2022). https://doi.org/10.3389/fpsyg.2022.815665.
17. Knierim, M.T., Rissler, R., Dorner, V., Maedche, A., Weinhardt, C.: The psychophysiology of flow: A systematic review of peripheral nervous system features. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., and Randolph, A.B. (eds.) Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017. pp. 109–120. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67431-5_13.

18. Stangl, F.J., Kalischko, T.: Unpacking the flow experience: A thematic analysis of neurophysiological flow research. In: Proceedings of the 21st International Conference on Wirtschaftsinformatik (2026).
19. Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: A typology of literature reviews. Information & Management. 52, 183–199 (2015). https://doi.org/10.1016/j.im.2014.08.008.
20. Schryen, G., Benlian, A., Rowe, F., Gregor, S., Larsen, K., Petter, S., Paré, G., Wagner, G., Haag, S., Yasasin, E.: Literature reviews in IS research: What can be learnt from the past and other fields? Communications of the Association for Information Systems. 41, 759–774 (2017). https://doi.org/10.17705/1CAIS.04130.
21. Schryen, G., Wagner, G., Benlian, A., Paré, G.: A knowledge development perspective on literature reviews: Validation of a new typology in the IS field. Communications of the Association for Information Systems. 46, 134–186 (2020). https://doi.org/10.17705/1CAIS.04607.
22. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly. 26, xiii–xxiii (2002).
23. vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: Newell, S., Whitley, E.A., Pouloudi, N., Wareham, J., and Mathiassen, L. (eds.) Proceedings of the 17th European Conference on Information Systems. pp. 2206–2217 (2009).
24. Riedl, R., Fischer, T., Léger, P.-M., Davis, F.D.: A decade of NeuroIS research: Progress, challenges, and future directions. ACM SIGMIS Database: The DATA BASE for Advances in Information Systems. 51, 13–54 (2020). https://doi.org/10.1145/3410977.3410980.
25. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. eds: Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2015. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18702-0.
26. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. eds: Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2016. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-41402-7.
27. 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 2025, Vienna, Austria. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-00815-2.
28. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. eds: Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67431-5.
29. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. eds: Information Systems and Neuroscience: NeuroIS Retreat 2018. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01087-4.
30. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Fischer, T. eds: Information Systems and Neuroscience: NeuroIS Retreat 2019. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-28144-1.
31. Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Fischer, T. eds: Information Systems and Neuroscience: NeuroIS Retreat 2020. Springer, Cham (2020).

https://doi.org/10.1007/978-3-030-60073-0.
32. 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 2021. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88900-5.
33. 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. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13064-9.
34. 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 2023, Vienna, Austria. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-58396-4.
35. 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 2024, Vienna, Austria. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9.
36. Kivikangas, J.M.: Psychophysiology of flow experience: An explorative study. , Master's Thesis. University of Helsinki (2006).
37. Sajno, E.: Affective Computing for detecting psychological Flow state: A definition and methodological problem. In: Proceedings of the 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos. pp. 1–5 (2023). https://doi.org/10.1109/ACIIW59127.2023.10388109.
38. Li, Z., Qiao, K., Zou, Z.: A measurement system for assessing interaction levels with intelligent chatbots: The flow state measurement system. IEEE Instrumentation & Measurement Magazine. 27, 23–29 (2024). https://doi.org/10.1109/MIM.2024.10743134.
39. Ahmed, Y., Ferguson-Pell, M., Adams, K., Ríos Rincón, A.: EEG-based engagement monitoring in cognitive games. Sensors. 25, 2072 (2025). https://doi.org/10.3390/s25072072.
40. Bigne, E., Ruiz, C., Curras-Perez, R.: Furnishing your home? The impact of voice assistant avatars in virtual reality shopping: A neurophysiological study. Computers in Human Behavior. 153, 108104 (2024). https://doi.org/10.1016/j.chb.2023.108104.
41. Bishop, A., MacNeil, E., Izzetoglu, K.: Cognitive workload quantified by physiological sensors in realistic immersive settings. In: Schmorrow, D.D. and Fidopiastis, C.M. (eds.) Augmented Cognition: 15th International Conference, AC 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings. pp. 119–133. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78114-9_9.
42. Guo, J., Zhao, Y., Zhang, W., Lu, K., Feng, X., Xia, T.: The influence of the information richness of interfaces on consumers' purchase intention: The sequential mediating effects of cognitive load, mental imagery, and flow experience. Behavioral Sciences. 15, 673 (2025). https://doi.org/10.3390/bs15050673.
43. Klarkowski, M., Johnson, D., Wyeth, P., Phillips, C., Smith, S.: Psychophysiology of challenge in play: EDA and self-reported arousal. In: Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1930–1936 (2016). https://doi.org/10.1145/2851581.2892485.
44. Loudon, G., Zampelis, D., Deininger, G.: Using real-time biofeedback of heart rate variability measures to track and help improve levels of attention and relaxation. In: Proceedings of the ACM SIGCHI Conference on Creativity and Cognition. pp. 348–355 (2017). https://doi.org/10.1145/3059454.3059466.

45. de Manzano, Ö., Theorell, T., Harmat, L., Ullén, F.: The psychophysiology of flow during piano playing. Emotion. 10, 301–311 (2010). https://doi.org/10.1037/a0018432.
46. Schlagowski, R., Mertes, S., Schiller, D., Can, Y.S., André, E.: Exploring physiology-based classification of flow during musical improvisation in mixed reality. In: Fred, A., Hadjali, A., Gusikhin, O., and Sansone, C. (eds.) Deep Learning Theory and Applications: 5th International Conference, DeLTA 2024, Dijon, France, July 10–11, 2024, Proceedings, Part I. pp. 296–309. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-66694-0_18.
47. Léger, P.-M., Davis, F.D., Cronan, T.P., Perret, J.: Neurophysiological correlates of cognitive absorption in an enactive training context. Computers in Human Behavior. 34, 273–283 (2014). https://doi.org/10.1016/j.chb.2014.02.011.
48. Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., Sun, X., Meng, X.: A framework for physiological indicators of flow in VR games: Construction and preliminary evaluation. Personal and Ubiquitous Computing. 20, 821–832 (2016). https://doi.org/10.1007/s00779-016-0953-5.
49. Bian, Y., Yang, C., Gao, F., Li, H., Sun, X., Meng, X., Wang, Y.: A physiological evaluation model for flow-experience in VR games: Construction and preliminary test. In: Proceedings of the International Conference on Identification, Information, and Knowledge in the Internet of Things. pp. 244–249 (2016). https://doi.org/10.1109/IIKI.2015.59.
50. Drachen, A., Nacke, L.E., Yannakakis, G., Pedersen, A.L.: Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In: Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games. pp. 49–54 (2010). https://doi.org/10.1145/1836135.1836143.
51. Drachen, A., Nacke, L.E., Yannakakis, G., Pedersen, A.L.: Psychophysiological correlations with gameplay experience dimensions. In: Workshop Proceedings of Brain, Body and Bytes at the ACM Conference on Human Factors in Computing Systems (2010).
52. Nacke, L., Lindley, C.A.: Flow and immersion in first-person shooters: measuring the player's gameplay experience. In: Proceedings of the Conference on Future Play: Research, Play, Share. pp. 81–88 (2008). https://doi.org/10.1145/1496984.1496998.
53. Nacke, L.E., Lindley, C.A.: Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. Loading...: The Journal of the Canadian Game Studies Association. 3, 1–21 (2009).
54. Stangl, F.J., Riedl, R.: Neurophysiological measurements in the research field of interruption science: Insights into applied methods for different interruption types based on an umbrella review. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2023, Vienna, Austria. pp. 123–152. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-58396-4_11.
55. Stangl, F.J., Riedl, R.: Autonomic nervous system activity measurements in the research field of interruption science: Insights into applied methods. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2024, Vienna, Austria. pp. 47–69. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9_6.
56. Stangl, F.J., Riedl, R.: Brain imaging methods in the research field of interruption science: An analysis of applied methods. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M.,

Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2024, Vienna, Austria. pp. 71–93. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9_7.

57. Stangl, F.J., Riedl, R.: Hormone measurements in the research field of interruption science: Review of applied methods. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2024, Vienna, Austria. pp. 95–114. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-71385-9_8.
58. Chin, M.S., Kales, S.N.: Is there an optimal autonomic state for enhanced flow and executive task performance? Frontiers in Psychology. 10, 1716 (2019). https://doi.org/10.3389/fpsyg.2019.01716.
59. Gaggioli Andrea, Cipresso Pietro, Serino Silvia, Riva Giuseppe: Psychophysiological correlates of flow during daily activities. In: Wiederhold, B.K. and Riva, G. (eds.) Annual Review of Cybertherapy and Telemedicine 2013. pp. 65–69. IOS Press, Fairfax (2013). https://doi.org/10.3233/978-1-61499-282-0-65.
60. Khanganba, S.P., Ubaradka, A.: Immersion in the shadows: Physiological markers of flow experience in a virtual reality survival horror game. In: Jin, S., Kim, J.H., Kong, Y.-K., Park, J., and Yun, M.H. (eds.) Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 4: Better Life Ergonomics for Future Humans (IEA 2024). pp. 454–460. Springer, Singapore (2025). https://doi.org/10.1007/978-981-95-0289-9_67.
61. Lemmens, J.S., Clark, F.E., Lyu, X., Taylor, A.: Time flies when you're having flow: An experiment on time perception and challenge in a VR game. Computers in Human Behavior Reports. 18, 100664 (2025). https://doi.org/10.1016/j.chbr.2025.100664.
62. Sajno, E., Rossi, A., Beretta, A., Novielli, N., Pappalardo, L., Riva, G.: What tugs at your heartstrings? Exploring flow, and affect recognition through HRV, while video gaming. In: Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos. pp. 214–217 (2024). https://doi.org/10.1109/ACIIW63320.2024.00045.
63. Tozman, T., Magdas, E.S., MacDougall, H.G., Vollmeyer, R.: 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 (2015). https://doi.org/10.1016/j.chb.2015.06.023.
64. Knierim, M.T., Rissler, R., Hariharan, A., Nadj, M., Weinhardt, C.: Exploring flow psychophysiology in knowledge work. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., and Randolph, A.B. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2018. pp. 239–249. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01087-4_29.
65. Knierim, M.T., Pieper, V., Schemmer, M., Loewe, N., Reali, P.: Predicting in-field flow experiences over two weeks from ECG data: A case study. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2021. pp. 96–102. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88900-5_11.
66. Prajod, P., Lavit Nicora, M., Mondellini, M., Falerni, M.M., Vertechy, R., Malosio, M., André, E.: Flow in human-robot collaboration—multimodal analysis and perceived

challenge detection in industrial scenarios. Frontiers in Robotics and AI. 11, 1393795 (2024). https://doi.org/10.3389/frobt.2024.1393795.

67. Nacke, L.E., Grimshaw, M.N., Lindley, C.A.: More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. Interacting with Computers. 22, 336–343 (2010). https://doi.org/10.1016/j.intcom.2010.04.005.
68. Ye, X., Ning, H., Backlund, P., Ding, J.: Flow experience detection and analysis for game users by wearable-devices-based physiological responses capture. IEEE Internet of Things Journal. 8, 1373–1387 (2021). https://doi.org/10.1109/JIOT.2020.3010853.
69. Bastarache-Roberge, M.-C., Léger, P.-M., Courtemanche, F., Sénécal, S., Fredette, M.: Measuring flow using psychophysiological data in a multiplayer gaming context. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., and Randolph, A.B. (eds.) Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2015. pp. 187–191. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-18702-0_25.
70. Bian, Y., Zhou, C., Gai, W., Liu, J., Yang, C.: The effect of embodied interaction designs on flow experience: Examination in VR games. Virtual Reality. 27, 1549–1565 (2023). https://doi.org/10.1007/s10055-023-00758-3.
71. Tian, Y., Bian, Y., Han, P., Wang, P., Gao, F., Chen, Y.: Physiological signal analysis for evaluating flow during playing of computer games of varying difficulty. Frontiers in Psychology. 8, 1121 (2017). https://doi.org/10.3389/fpsyg.2017.01121.
72. Qi, W., Zhang, D., Nourrit, V., Bougrenet, J.-L. de, Ye, L.: Real-time flow state analysis in game: A non-intrusive system based on eye feature detection. In: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. pp. 1508–1509 (2025). https://doi.org/10.1109/VRW66409.2025.00400.
73. Harris, D.J., Vine, S.J., Wilson, M.R.: Is flow really effortless? The complex role of effortful attention. Sport, Exercise, and Performance Psychology. 6, 103–114 (2017). https://doi.org/10.1037/spy0000083.
74. Wang, Y., Chen, B., Liu, H., Hu, Z.: Understanding flow experience in video learning by multimodal data. International Journal of Human–Computer Interaction. 40, 3144–3158 (2024). https://doi.org/10.1080/10447318.2023.2181878.
75. Bartholomeyczik, K., Knierim, M.T., Nieken, P., Seitz, J., Stano, F., Weinhardt, C.: Flow in knowledge work: An initial evaluation of flow psychophysiology across three cognitive tasks. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2022. pp. 23–33. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13064-9_3.
76. Beiramvand, M., Koivula, R., Lipping, T.: Development of an EEG-based method for detecting flow state using a wearable headband in a game environment. In: Proceedings of the 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1–6 (2025). https://doi.org/10.1109/EMBC58623.2025.11251885.
77. Lin, B., Guo, B., Zhuang, L., Zhang, D., Wang, F.: Neural oscillations predict flow experience. Cognitive Neurodynamics. 19, 1 (2025). https://doi.org/10.1007/s11571-024-10205-x.
78. Lu, H., Van Der Linden, D., Bakker, A.B.: The neuroscientific basis of flow: Learning progress guides task engagement and cognitive control. NeuroImage. 308, 121076 (2025). https://doi.org/10.1016/j.neuroimage.2025.121076.
79. McMahan, T., Parberry, I., Parsons, T.D.: Evaluating player task engagement and arousal

using electroencephalography. Procedia Manufacturing. 3, 2303–2310 (2015). https://doi.org/10.1016/j.promfg.2015.07.376.
80. Meng, L., Pei, G., Zheng, J., Ma, Q.: Close games versus blowouts: Optimal challenge reinforces one's intrinsic motivation to win. International Journal of Psychophysiology. 110, 102–108 (2016). https://doi.org/10.1016/j.ijpsycho.2016.11.001.
81. Nacke, L.E., Stellmach, S., Lindley, C.A.: Electroencephalographic assessment of player experience: A pilot study in affective ludology. Simulation & Gaming. 42, 632–655 (2011). https://doi.org/10.1177/1046878110378140.
82. Omezzine, M.A.: Keeping Learners in the Flow: How AI-Driven Personalization Balances Cognitive Load and Enhances Performance. In: Americas Conference on Information Systems, AMCIS 2025. pp. 2447–2451. Association for Information Systems (2025).
83. Plotnikov, A., Stakheika, N., De Gloria, A., Schatten, C., Bellotti, F., Berta, R., Fiorini, C., Ansovini, F.: Exploiting real-time EEG analysis for assessing flow in games. In: Proceedings of the IEEE 12th International Conference on Advanced Learning Technologies. pp. 688–689 (2012). https://doi.org/10.1109/ICALT.2012.144.
84. Shilong, L., Tong, C., Liu, Z., Li, X., Zheng, Y., Zhou, C., Liu, J., Bian, Y.: Assessing dynamic flow experience from EEG signals: A processing-based approach. In: Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology. pp. 1–19 (2025). https://doi.org/10.1145/3746059.3747603.
85. Wang, C.-C., Hsu, M.-C.: An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction. Information & Management. 51, 912–923 (2014). https://doi.org/10.1016/j.im.2014.05.010.
86. Wang, C.-C., Hsu, M.-C.: Flow experience and challenge-skill balance in e-learning. In: Proceedings of the 17th Pacific Asia Conference on Information Systems (2013).
87. Chanel, G., Rebetez, C., Bétrancourt, M., Pun, T.: Emotion assessment from physiological signals for adaptation of game difficulty. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 41, 1052–1063 (2011). https://doi.org/10.1109/TSMCA.2011.2116000.
88. Wu, S.-F., Lu, Y.-L., Lien, C.-J.: Detecting students' flow states and their construct through electroencephalogram: Reflective flow experiences, balance of challenge and skill, and sense of control. Journal of Educational Computing Research. 58, 1515–1540 (2021). https://doi.org/10.1177/0735633120944084.
89. Hang, Y., Unenbat, B., Tang, S., Wang, F., Lin, B., Zhang, D.: Exploring the neural correlates of flow experience with multifaceted tasks and a single-channel prefrontal EEG recording. Sensors. 24, 1894 (2024). https://doi.org/10.3390/s24061894.
90. Katahira, K., Yamazaki, Y., Yamaoka, C., Ozaki, H., Nakagawa, S., Nagata, N.: EEG correlates of the flow state: A combination of increased frontal theta and moderate frontocentral alpha rhythm in the mental arithmetic task. Frontiers in Psychology. 9, 300 (2018). https://doi.org/10.3389/fpsyg.2018.00300.
91. Knierim, M.T., Stano, F., Kurz, F., Heusch, A., Wilson, M.L.: Exploring flow in real-world knowledge work using discrete cEEGrid sensors. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–16 (2025). https://doi.org/10.1145/3706598.3713512.
92. Knierim, M.T., Nadj, M., Hariharan, A., Weinhardt, C.: Flow neurophysiology in knowledge work: Electroencephalographic observations from two cognitive tasks. In:

Belda-Lois, J.-M., Wang, C., Morales, M.J.D., Pope, A., and da Silva, H.P. (eds.) Proceedings of the 5th International Conference on Physiological Computing Systems. pp. 42–53 (2018). https://doi.org/10.5220/0006926700420053.

93. Knierim, M.T., Zimny, C., Seitz, J.: Monitoring cognitive load with 3D-printed EEG headphones. In: Proceedings of the 43rd International Conference on Information Systems (2022).
94. Labonté-LeMoyne, É., Léger, P.-M., Resseguier, B., Bastarache-Roberge, M.-C., Fredette, M., Sénécal, S., Courtemanche, F.: Are we in flow neurophysiological correlates of flow states in a collaborative game. In: Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems. pp. 1980–1988 (2016). https://doi.org/10.1145/2851581.2892356.
95. Liang, T.-P., Lin, Y.-L., Shiau, W.-L., Chen, S.-F.: Investigating common method bias via an EEG study of the flow experience in website design. Journal of Electronic Commerce Research. 22, 305–321 (2021).
96. Klasen, M., Weber, R., Kircher, T.T.J., Mathiak, K.A., Mathiak, K.: Neural contributions to flow experience during video game playing. Social Cognitive and Affective Neuroscience. 7, 485–495 (2012). https://doi.org/10.1093/scan/nsr021.
97. Ulrich, M., Keller, J., Grön, G.: Dorsal raphe nucleus down-regulates medial prefrontal cortex during experience of flow. Frontiers in Behavioral Neuroscience. 10, 169 (2016). https://doi.org/10.3389/fnbeh.2016.00169.
98. Ulrich, M., Keller, J., Hoenig, K., Waller, C., Grön, G.: Neural correlates of experimentally induced flow experiences. NeuroImage. 86, 194–202 (2014). https://doi.org/10.1016/j.neuroimage.2013.08.019.
99. Nozawa, T., Kondo, M., Yamamoto, R., Jeong, H., Ikeda, S., Sakaki, K., Miyake, Y., Ishikawa, Y., Kawashima, R.: Prefrontal inter-brain synchronization reflects convergence and divergence of flow dynamics in collaborative learning: A pilot study. Frontiers in Neuroergonomics. 2, 686596 (2021). https://doi.org/10.3389/fnrgo.2021.686596.
100. Yoshida, K., Sawamura, D., Inagaki, Y., Ogawa, K., Ikoma, K., Sakai, S.: Brain activity during the flow experience: A functional near-infrared spectroscopy study. Neuroscience Letters. 573, 30–34 (2014). https://doi.org/10.1016/j.neulet.2014.05.011.
101. Yu, D., Wang, S., Song, F., Liu, Y., Zhang, S., Wang, Y., Xie, X., Zhang, Z.: Research on user experience of the video game difficulty based on flow theory and fNIRS. Behaviour & Information Technology. 42, 789–805 (2023). https://doi.org/10.1080/0144929X.2022.2043442.
102. Watanabe, S., Suzuki, T., Ono, Y.: Predicting the Intensity of the Flow State Using EEG and fNIRS Biomarkers. Advanced Biomedical Engineering. 13, 335–342 (2024). https://doi.org/10.14326/abe.13.335.
103. de Manzano, Ö., Cervenka, S., Jucaite, A., Hellenäs, O., Farde, L., Ullén, F.: Individual differences in the proneness to have flow experiences are linked to dopamine D2-receptor availability in the dorsal striatum. NeuroImage. 67, 1–6 (2013). https://doi.org/10.1016/j.neuroimage.2012.10.072.
104. Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., Glenk, L.M.: Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. Computers & Education. 79, 69–100 (2014). https://doi.org/10.1016/j.compedu.2014.07.001.

105. Peifer, C., Schächinger, H., Engeser, S., Antoni, C.H.: Cortisol effects on flow-experience. Psychopharmacology. 232, 1165–1173 (2015). https://doi.org/10.1007/s00213-014-3753-5.
106. de Sampaio Barros, M.F., Araújo-Moreira, F.M., Trevelin, L.C., Radel, R.: Flow experience and the mobilization of attentional resources. Cognitive, Affective, & Behavioral Neuroscience. 18, 810–823 (2018). https://doi.org/10.3758/s13415-018-0606-4.
107. Harmat, L., de Manzano, Ö., Theorell, T., Högman, L., Fischer, H., Ullén, F.: Physiological correlates of the flow experience during computer game playing. International Journal of Psychophysiology. 97, 1–7 (2015). https://doi.org/10.1016/j.ijpsycho.2015.05.001.
108. Rácz, M., Becske, M., Magyaródi, T., Kitta, G., Szuromi, M., Márton, G.: Physiological assessment of the psychological flow state using wearable devices. Scientific Reports. 15, 11839 (2025). https://doi.org/10.1038/s41598-025-95647-x.
109. Mauri, M., Cipresso, P., Balgera, A., Villamira, M., Riva, G.: Why is Facebook so successful? Psychophysiological measures describe a core flow state while using Facebook. Cyberpsychology, Behavior, and Social Networking. 14, 723–731 (2011). https://doi.org/10.1089/cyber.2010.0377.
110. Berta, R., Bellotti, F., De Gloria, A., Pranantha, D., Schatten, C.: Electroencephalogram and physiological signal analysis for assessing flow in games. IEEE Transactions on Computational Intelligence and AI in Games. 5, 164–175 (2013). https://doi.org/10.1109/TCIAIG.2013.2260340.
111. Ulrich, M., Keller, J., Grön, G.: Neural signatures of experimentally induced flow experiences identified in a typical fMRI block design with BOLD imaging. Social Cognitive and Affective Neuroscience. 11, 496–507 (2016). https://doi.org/10.1093/scan/nsv133.
112. Bieri, U., Baumgartner, M., Richoz, A.-R., Eberli, D., Do, C.T., Poyet, C.: Investigation of flow state occurrence during robotic virtual reality operations. Scientific Reports. 15, 40710 (2025). https://doi.org/10.1038/s41598-025-24215-0.
113. Khoshnoud, S., Alvarez Igarzábal, F., Wittmann, M.: Brain–heart interaction and the experience of flow while playing a video game. Frontiers in Human Neuroscience. 16, 819834 (2022). https://doi.org/10.3389/fnhum.2022.819834.
114. Keller, J., Bless, H., Blomann, F., Kleinböhl, D.: Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. Journal of Experimental Social Psychology. 47, 849–852 (2011). https://doi.org/10.1016/j.jesp.2011.02.004.
115. Peifer, C., Schulz, A., Schächinger, H., Baumann, N., Antoni, C.H.: The relation of flow-experience and physiological arousal under stress — Can u shape it? Journal of Experimental Social Psychology. 53, 62–69 (2014). https://doi.org/10.1016/j.jesp.2014.01.009.
116. Balapour, A., Riedl, R.: Ecological validity in NeuroIS research: Theory, evidence, and a roadmap for future studies. Journal of the Association for Information Systems. 26, 9–65 (2025). https://doi.org/10.17705/1jais.00904.
117. Kalischko, T., Riedl, R.: Physiological measurement in the research field of electronic performance monitoring: Review and a call for NeuroIS studies. 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 2020. pp. 233–243. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60073-0_27.
118. Riedl, R., Maedche, A.: Why behavioral measures as indirect measures of neurophysiological activity are important for progress in the NeuroIS discipline. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2025, Vienna, Austria. pp. 159–167. Springer, Cham (2025). https://doi.org/10.1007/978-3-032-00815-2_15.
119. Stangl, F.J., Riedl, R.: 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., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2022. pp. 269–283. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13064-9_28.
120. Stangl, F.J., Riedl, R.: 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., and Müller-Putz, G.R. (eds.) Information Systems and Neuroscience: NeuroIS Retreat 2022. pp. 285–299. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-13064-9_29.
121. Stangl, F.J., Riedl, R.: Zum Erkenntnispotenzial kardiovaskulärer Messungen in der Wirtschaftsinformatik: Grundlagen und Ergebnisse empirischer Forschung. In: Schütte, R., Hohmann, S., Krol, B., and Peters, M.L. (eds.) Produktions- und Informationsmanagement: Festschrift für Prof. Dr. Stephan Zelewski. pp. 361–387. Springer, Wiesbaden (2024). https://doi.org/10.1007/978-3-658-46113-3_15.
122. Riedl, R., Davis, F.D., Hevner, A.R.: Towards a NeuroIS research methodology: Intensifying the discussion on methods, tools, and measurement. Journal of the Association for Information Systems. 15, I–XXXV (2014). https://doi.org/10.17705/1jais.00377.

Toward Multimodal Attention Detection in Virtual Reality Lectures: Combining Ear-EEG and Eye Tracking

Fabio Stano[1,*], and Thimo Schulz[1]

[1] Karlsruhe Institute of Technology (KIT), Institute for Information Systems (WIN), Karlsruhe, Germany
{fabio.stano,thimo.schulz}@kit.edu

Abstract. Virtual reality (VR) learning environments enable new forms of adaptive education but require reliable sensing of learners' attentional states. In this study, we explore the feasibility of multimodal attention tracking in VR lectures by combining eye tracking with ear-centered EEG (ear-EEG). Participants attended VR lectures while gaze behavior and neural signals were recorded. Visual attention was analyzed using gaze allocation to task-relevant regions, while auditory attention was examined using auditory steady-state response and neural speech envelope tracking. Results indicate that gaze significantly predicts noticing of lecture content, whereas auditory attention detection remains technically challenging with ear-EEG. Survey and qualitative responses indicate cautious user acceptance under strong privacy safeguards. The findings provide initial insights into multimodal attention tracking for adaptive immersive learning systems.

Keywords: Virtual Reality (VR) · Education · Eye Tracking · Auditory Attention Detection (AAD) · cEEGrids

Introduction

Immersive virtual reality (VR) learning environments offer new opportunities for adaptive and personalized education, but such adaptation requires reliable information about learners' internal states during instruction [1, 2]. One particularly important state is attention, as effective learning depends on attending to relevant content. While visual attention in VR can be readily captured through integrated eye tracking, providing a direct measure of gaze behavior [3, 4], attention in lecture settings is not limited to the visual channel. A substantial portion of instructional content is delivered through speech, making auditory attention equally important, yet inherently more difficult to assess because it is not directly observable [5].

To address this challenge, NeuroIS research increasingly draws on electroencephalography (EEG), including approaches such as Auditory Steady-State Response (ASSR) tracking and neural speech envelope tracking (NET), to infer how strongly learners process auditory input [6, 7]. Applying these methods in immersive VR, however, introduces practical constraints, as neural sensing must remain compatible with head-mounted displays and realistic interaction. Wearable solutions such as ear-centered EEG offer a promising alternative to traditional full-cap systems [8]. Against this background, this work explores the feasibility of multimodal attention tracking in VR lectures by combining eye tracking and around-the-ear-EEG. To investigate this idea, we pose the following research questions:

RQ1: To what extent can visual and auditory attention states in VR lectures be detected using ear-EEG and eye tracking?

RQ2: Does multimodal tracking outperform unimodal attention detection approaches?

RQ3: How do users perceive multimodal attention detection in immersive educational scenarios?

To answer these questions, we plan a laboratory experiment in which participants attend VR lectures under controlled attention conditions while their eye movements and ear-EEG signals are recorded. In a first pilot study (n = 6), we primarily assess the technical feasibility of detecting visual and auditory attention signals and explore participants' perceptions of such an attention-aware system. While the overall study design targets all three research questions, the present research-in-progress stage focuses on providing initial insights into RQ1 and RQ3, whereas a robust evaluation of multimodal performance advantages (RQ2) is reserved for future work with a larger sample.

Theoretical Background and Related Work

Visual attention tracking is a well-established NeuroIS stream for studying how users allocate limited cognitive resources in digital environments. Eye tracking is central to this work because it captures overt visual attention [2, 3]. Prior IS research shows that gaze data can reveal information-processing strategies beyond self-reports and behavioral outcomes [9]. In immersive VR, this becomes especially relevant because environments are dynamic, spatial, and cognitively demanding. Accordingly, recent reviews identify eye tracking as a key method for studying attention in VR [4]. At the same time, gaze alone may not fully capture latent attentional states, motivating the integration of EEG and eye tracking in extended reality (XR) and VR research [10].

Two of the most established methods for tracking **auditory attention** are Auditory Steady State Response tracking and Neural Envelope Tracking.

ASSRs are frequency-specific neural responses elicited by periodically amplitude-modulated sounds [11]. By tagging auditory streams with distinct modulation frequencies, stimulus-specific neural activity can be isolated in the EEG spectrum [12]. Selective attention modulates ASSR amplitude, typically enhancing responses to attended streams [13]. However, attentional effects depend on modulation frequency and task design [6]. Additionally, ASSR amplitudes are relatively small, require sufficient signal-to-noise ratio and averaging, and may saturate or be reduced when speech-like properties are introduced compared to pure tones or chirps [14].

NET works because continuous speech contains slow amplitude fluctuations (1–8 Hz) that are tracked by cortical activity [15]. Neural envelope tracking is commonly quantified using encoding or decoding models that relate the speech envelope to EEG activity, for example through temporal response functions (TRFs) estimated with linear regression models . Attention strengthens the coupling between the speech envelope and neural activity, resulting in stronger tracking of attended speech relative to ignored signals [16, 17]. Recent work increasingly explores nonlinear decoding approaches, including deep neural networks, to improve attention classification performance in more complex listening environments [7].

Multimodal attention tracking in virtual reality seeks to capture attentional processes more comprehensively by combining complementary data streams such as EEG, eye tracking, and, in some cases, head-movement measures. Early work by Delvigne et al. [18] showed that attention in VR can be estimated by integrating EEG with gaze direction, pupil diameter, and head movement, pointing to the value of multimodal inference beyond single-sensor approaches. More recent studies have applied this approach in increasingly naturalistic settings. For example, Huizeling et al. [19] combined EEG and 3D eye tracking to study anticipatory processing during speech comprehension in virtual environments, while Levy et al. [20] used EEG and gaze data in a VR classroom to examine selective attention to a teacher's speech under auditory distraction. These studies showcase the feasibility of multimodal attention tracking in VR, while leaving the question unanswered how such attention tracking relates to actual knowledge absorption.

Methodology

Our **pilot study design** was implemented to assess the feasibility of multimodal attention measurement in a VR lecture setting. First, skin preparation and gel application were performed, with electrode impedances targeted below 30 kOhm. Setup followed a fixed sequence: (1) electrode placement, (2) fitting and comfort adjustment of the VR headset, (3) amplifier attachment, and (4) signal quality checking. The following VR experiment was predefined and guided by the virtual tutor. It began with two resting-state conditions (90 s eyes closed 90 s eyes open with a fixation cross), followed by instructed lecture conditions of four minutes each: *listening* only, *reading* only (with audible but unintelligible foreign-language speech), and a *none* condition with the same foreign-language speech but instructions to not attend to anything lecture-related. The design choice to use foreign-language speech in the *reading* and *none* conditions aimed to reduce involuntary attentional capture by semantic content, thereby making it easier for participants to disengage from the auditory stream when instructed. Participants then completed a naturalistic 12-minute lecture phase with combined visual and auditory stimuli. After the VR session, the headset and EEG sensors were removed. Finally, participants completed a survey (implemented via oTree, [21]) covering demographics, a post-hoc recognition test on lecture statements, and general attitudes towards the concept and experiences with the experimental design. In line with the pilot nature of the study, the focus of the current implementation is on technical viability, user comfort, and initial signal quality rather than confirmatory hypothesis testing.

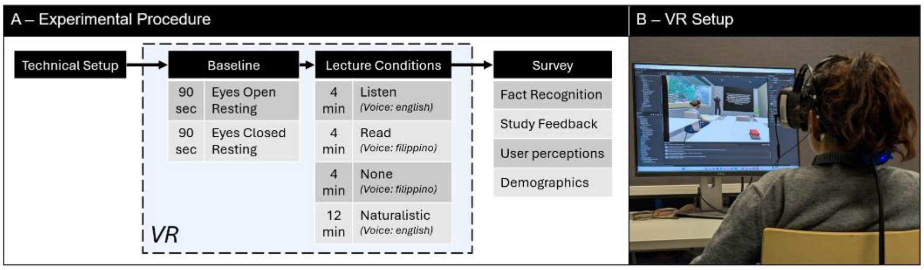


Fig. 11. Experimental Procedure (A) and VR Setup (B)

The **VR environment** was implemented in Unity 2022.3.30, building on the classroom design proposed by Liu et al. [22] To introduce controlled visual distraction, a fish tank was added to the scene. The setup used a Varjo XR-3 headset to ensure high-resolution, state-of-the-art VR presentation. Hand tracking was enabled to support interaction and increase immersion.

Lecture stimuli comprised six 4-minute VR lectures. Each lecture used black slides with white text, narrated by a virtual tutor using an AI-generated gender-neutral voice. The lecture contents covered six philosophical topics, and content order was randomized across participants. For ASSR tracking, we modulated the amplitude of the voice stimuli with 50% modulation depth, trying to strike a balance between keeping the voice natural, and improving ASSR signal-to-noise-ratio (SNR).

For the **EEG setup** we followed prior work [23], using gold-plated open-cEEGrid electrodes in combination with the low-cost, open-source OpenBCI Cyton amplifier. EEG was sampled across 8 channels at 250 Hz and recorded via the OpenBCI GUI. The amplifier was fixed to the headset using adhesive tape.

EEG data were processed and analyzed using MNE Python [24]. General preprocessing was kept minimal, using a 0.5-45Hz bandpass filter and visual bad channel inspection, resulting in no exclusions across all participants.

Qualitative responses were analyzed by the two authors using a low-inference content analysis approach, focusing on recurring themes across participants' statements.

Overall, **6 participants** (3 female, 3 male) from a convenience sample took part in the pilot study. For P1 the EEG recordings failed due to a technical issue, resulting in EEG data from 5 participants.

Preliminary Results

Visual Attention Analysis. Greater visual attention to task-relevant AOIs (Slides + Tutor) significantly predicted noticing (b = 1.43, SE = 0.45, p = .001, OR = 4.19). Each additional 10% of viewing time on task-relevant AOIs was associated with about 15% higher odds of noticing. When separated, only time on Slides remained significant (p = .004), whereas time on Tutor did not (p = .897). However, the model explained little variance (pseudo-R^2 = .034), suggesting that gaze alone captures only part of the noticing process.

Auditory Attention Analysis. To assess the feasibility of auditory attention detection in VR using ear-EEG, we first tested for the presence of a 40 Hz ASSR. After enhancing activity around 40 Hz with a generalized eigenvalue decomposition (GEVD) spatial filter, spectral power in overlapping 15 s windows was compared against neighboring frequency bins (±6 Hz) using an F-test ($\alpha = 0.01$), with low-energy stimulus periods excluded. Overall, ASSR detections were rare. For four participants (P2, P3, P5, P6), detection rates were near zero across conditions (0–7% of analyzed windows). In contrast, participant P4 showed robust ASSR detection across all conditions, ranging from 46% to 85% of valid windows, with the highest rate in the *open* condition. Shorter window sizes resulted in lower detection rates, while longer windows did not increase it further.

Speech envelope tracking. To examine whether neural tracking of speech differed between attention conditions, we trained a linear backward model on the first half of the *listen* condition and evaluated envelope reconstruction on the remaining *listen* data as well as on the *read* and *none* conditions, adapting established methodology [25]. Reconstruction accuracy was quantified as the Pearson correlation between the predicted and actual speech envelope within 30 s windows and averaged per participant and condition. Across participants, correlations were generally small (typical range −0.04 to 0.02), which is consistent with prior work on envelope tracking using ear-centered EEG sensors. Because correlation signs can vary depending on model polarity, we examined absolute correlations as a measure of reconstruction strength. Using this metric, envelope tracking was highest during the *listen* condition ($|r| = 0.016$), compared to the *read* ($|r| = 0.009$) and *none* conditions ($|r| = 0.007$).

Perception of the System. On 7-point Likert scales, participants evaluated the VR lecture setting positively, rating the lecture as interesting ($M = 5.50$), understandable ($M = 6.17$), and the tutor as relatively natural ($M = 5.50$); self-reported attention to the lecture was likewise high ($M = 5.50$). While the virtual environment was perceived as moderately realistic, distraction by the environment itself was low, whereas distraction by participants' own thoughts was comparatively higher, suggesting that attentional fluctuations may be driven more by internal processes than by salient visual elements. Acceptance judgments were strongest for video meetings ($M = 6.67$), online lectures ($M = 6.50$), and attention-critical jobs ($M = 6.00$), but clearly weaker for VR gaming ($M = 3.00$), indicating that multimodal attention detection is perceived as particularly relevant in educational and work-related settings, but less appropriate for entertainment scenarios.

Qualitative Insights. The system was largely perceived as promising for lecture contexts because it could *(a) support learners' attention in difficult presentations* and *(b) provide reflective feedback for targeted review*. However, acceptance hinges on strong privacy protections and clearly defined, bounded use: participants want the benefits without the feeling of being surveilled or exposing sensitive data.

Concerns cluster around *(a) privacy/security* and *(b) misuse of attention inference in asymmetric power contexts*, especially workplaces, where the system could become a surveillance or discipline tool. Secondary concerns include *data accuracy* (risk of wrong inferences) and *practical wearability* for extended use.

Participants primarily imagine the system as an *attention-aware assistant for learning and instruction*, extending from lectures to videos, demos, and reading. The *most powerful expansion lever is real-time detection*, which shifts the tool from retrospective reflection to immediate intervention and prompts suggestions for high-stakes domains (politics/military) that would require especially careful governance.

Discussion & Outlook

This pilot study examined the feasibility of multimodal attention tracking in immersive VR lectures using eye tracking and ear-EEG. The results highlight both the promise and the current limitations of combining these modalities in realistic learning environments.

Regarding **RQ1**, the findings indicate that visual attention provides a meaningful but incomplete proxy for learning-related processes. Visual attention to lecture-relevant regions significantly predicted noticing of lecture content, however only explaining a small portion of knowledge absorption. The relatively low explained variance suggests that gaze alone does not fully capture underlying cognitive processing. Learners may visually attend without deeply processing content, or conversely, disengage visually while remaining cognitively engaged with auditory input. This reinforces the need for complementary measures of attention beyond gaze.

For auditory attention, both ASSR and speech envelope tracking provided only limited evidence for robust detection in the current setup. ASSR responses were largely absent for most participants, with strong responses observed in only a single case. Similarly, speech envelope tracking yielded only weak reconstruction performance. Importantly, these results should not be interpreted as evidence against auditory attention tracking in VR, but rather as an indication of the methodological constraints of wearable, headset-compatible EEG systems. In particular, many attention-related neural responses are strongest in central and frontal regions, which are only indirectly captured by ear-centered EEG, limiting sensitivity in the present setup [26]. Several improvements for future work emerge from this pilot. First, longer recording periods, especially for training data, are likely necessary to stabilize neural decoding models. Second, increasing electrode coverage may improve spatial filtering and signal quality. Third, more advanced preprocessing and artifact handling will be critical in VR contexts, where movement-related noise is unavoidable. Finally, combining multiple neural features and leveraging nonlinear decoding approaches may help compensate for reduced signal quality.

Regarding **RQ2**, the pilot does not provide empirical evidence that multimodal tracking outperforms unimodal approaches. Instead, the results thus far support a conceptual rationale for multimodality: eye tracking and EEG capture complementary aspects of attention, namely overt allocation and underlying neural processing. The current findings therefore position multimodal attention tracking as a promising direction, but one that requires more robust sensing and integrated modeling before performance advantages can be systematically evaluated.

Regarding **RQ3**, participants perceived the system as promising, especially for lectures and other attention-critical contexts, but acceptance was clearly conditional. Qualitative feedback emphasized the value of such systems as supportive learning tools (for example, enabling reflection or identifying missed content) while also highlighting concerns around privacy, data security, inference accuracy, and potential misuse in asymmetric contexts such as workplaces. These findings underscore that attention-aware systems are not only a technical challenge but also a socio-technical one, requiring careful design and user-involvement to balance utility with autonomy and trust.

Taken together, the findings position multimodal attention tracking in VR as a promising but still methodologically demanding avenue for NeuroIS research. This pilot study advances NeuroIS research on attention-aware learning in VR by examining the feasibility and acceptance of multimodal attention tracking. Learning from these initial findings, the next step is to refine the sensing setup, experimental design and analysis pipeline, and test whether combining gaze and ear-EEG improves attention detection in immersive lecture scenarios. In addition, future studies should systematically assess user well-being factors such as VR-induced discomfort or cybersickness, which may influence both attention and data quality.

Beyond its methodological contribution, this work can inform the design of learner-centered VR systems that support attention while respecting privacy and responsible use.

References

1. Marougkas, A., Troussas, C., Krouska, A., Sgouropoulou, C.: How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. Multimed Tools Appl. 83, 18185–18233 (2023).
2. Dimoka, A., Banker, R., Benbasat, I., Davis, F., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Pavlou, P.A., Muller-Putz, G.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly. 36, (2012).
3. Vasseur, A., Passalacqua, M., Polytechnique Montréal, Sénécal, S., HEC Montréal, Léger, P.-M., HEC Montréal: The Use of Eye-tracking in Information Systems Research: A Literature Review of the Last Decade. THCI. 15, 292–321 (2023).
4. Shadiev, R., Li, D.: A review study on eye-tracking technology usage in immersive virtual reality learning environments. Computers & Education. 196, 104681 (2023).
5. Holtze, B., Rosenkranz, M., Jaeger, M., Debener, S., Mirkovic, B.: Ear-EEG Measures of Auditory Attention to Continuous Speech. Front. Neurosci. 16, 869426 (2022).
6. Linden, R.D., Picton, T.W., Hamel, G., Campbell, K.B.: Human auditory steady-state evoked potentials during selective attention. Electroencephalography and Clinical Neurophysiology. 66, 145–159 (1987).
7. Belo, J., Clerc, M., Schön, D.: EEG-Based Auditory Attention Detection and Its Possible Future Applications for Passive BCI. Front. Comput. Sci. 3, 661178 (2021).
8. Knierim, M.T., Puhl, D., Ivucic, G., Röddiger, T.: OpenBCI + 3D-Printed Headphones = Open ExG Headphones – An Open-Source Research Platform for Biopotential Earable Applications. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–7. ACM, Hamburg Germany (2023).
9. Bera, P., Soffer, P., Parsons, J.: Using Eye Tracking to Expose Cognitive Processes in Understanding Conceptual Models1. MIS Quarterly. 43, 1105–1126 (2019).

10. Long, X., Mayer, S., Chiossi, F.: Multimodal Detection of External and Internal Attention in Virtual Reality using EEG and Eye Tracking Features. In: Proceedings of Mensch und Computer 2024. pp. 29–43. ACM, Karlsruhe Germany (2024).
11. Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D.W.: Human auditory steady-state responses. International Journal of Audiology. (2003).
12. Manting, C.L., Gulyás, B., Ullén, F., Lundqvist, D.: Steady-state responses to concurrent melodies: source distribution, top-down, and bottom-up attention. Cerebral Cortex. (2022).
13. Mahajan, Y., Davis, C., Kim, J.: Attentional Modulation of Auditory Steady-State Responses. PLoS ONE. 9, e110902 (2014).
14. Sergeeva, A., Christensen, C.B., Kidmose, P.: Towards ASSR-based hearing assessment using natural sounds. J. Neural Eng. 21, 026045 (2024).
15. Kerlin, J.R., Shahin, A.J., Miller, L.M.: Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party." J. Neurosci. 30, 620–628 (2010).
16. Ding, N., Simon, J.Z.: Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. U.S.A. 109, 11854–11859 (2012).
17. Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J.: Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. Journal of Neurophysiology. 102, 349–359 (2009).
18. Delvigne, V., Wannous, H., Vandeborre, J.-P., Ris, L., Dutoit, T.: Attention Estimation in Virtual Reality with EEG based Image Regression. In: 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). pp. 10–16. IEEE, Utrecht, Netherlands (2020).
19. Huizeling, E., Alday, P.M., Peeters, D., Hagoort, P.: Combining EEG and 3D-eye-tracking to study the prediction of upcoming speech in naturalistic virtual environments: A proof of principle. Neuropsychologia. 191, 108730 (2023).
20. Levy, O., Libman Hackmon, S., Zvilichovsky, Y., Korisky, A., Bidet-Caulet, A., Schweitzer, J.B., Zion Golumbic, E.: Selective attention and sensitivity to auditory disturbances in a virtually real classroom. eLife. 13, RP103235 (2025).
21. Chen, D.L., Schonger, M., Wickens, C.: oTree—An open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance. 9, 88–97 (2016).
22. Liu, S., Feick, M., Bierhoff, L., Maedche, A.: AttentiveLearn: Personalized Post-Lecture Support for Gaze-Aware Immersive Learning, http://arxiv.org/abs/2603.05324, (2026).
23. Knierim, M.T., Stano, F., Kurz, F., Heusch, A., Wilson, M.L.: Exploring Flow in Real-World Knowledge Work Using Discrete cEEGrid Sensors. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. pp. 1–16. ACM, Yokohama Japan (2025).
24. Gramfort, A.: MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, (2013).
25. Mirkovic, B., Bleichner, M.G., De Vos, M., Debener, S.: Target Speaker Detection with Concealed EEG Around the Ear. Front. Neurosci. 10, (2016).
26. Wang, Y., Liu, X., Cui, H., Li, Z., Chen, X.: An auditory selective attention brain-computer interface system based on auditory steady-state response. Applied Acoustics. 228, 110291 (2025).

Panoptic Oversight and Critical Thinking in AI-Assisted Professional Decision-Making: A NeuroIS Work-in-Progress Study

Marion Korosec-Serfaty[1*], Amélie Lesvéque[2], Sylvain Sénécal[2], and Pierre-Majorique Léger[2]

[1] Department of Analytics, Operations and Information Technologies, School of Management, Université du Québec à Montréal (ESG UQAM), Montréal, Québec, Canada
marion.korosec-serfaty@uqam.ca

[2]Department of Information Technologies, HEC Montréal, Montréal, Québec, Canada
{amelie.lesveque, sylvain.senecal, pierre-majorique.leger}@hec.ca

Abstract. Critical thinking is a central safeguard for responsibility and accountability in deontologically-governed professions. Previous multi-method NeuroIS research has established how varying levels of AI reconstructive causal explanation completeness influence the physiological and behavioral dynamics of critical thinking in AI-assisted decision-making. This work-in-progress extends these findings by replicating the experimental methods under a panoptic manipulation reflecting the permanent possibility of professional audit. Preliminary findings, based on electrodermal activity, behavioral, and self-report measures of $N = 14$ participants, suggest that panopticism preserves deliberative-phase but reconfigures executive-phase dynamics of critical thinking, with the distinction lying not in explanation completeness but in the presence or absence of AI support. These preliminary observations suggest that institutional oversight practices designed for self-directed professional judgment may not straightforwardly extend to AI-assisted decision-making.

Keywords: Critical thinking · Panopticism · AI-assisted decision-making · Reconstructive AI explanations · Electrodermal activity

Introduction

Critical thinking is a central safeguard for responsibility and accountability in deontologically-governed professions [1, 2]. However, as artificial intelligence (AI)-assisted decision-making is increasingly integrated into these professional workflows, its inherent inscrutability constrains professionals' ability to critically evaluate and justify AI-informed decisions, increasing the risk of overreliance and potentially severing decisions from the deontological standards these professions are bound to uphold [3, 4].

In addressing this challenge, our prior NeuroIS research has shown that embedding AI explainability in the form of causal reconstructive explanations (post-hoc rationales that articulate decision-relevant factors in user-understandable, usable forms for judgment, hereinafter reconstructive explanations) influenced professionals' critical thinking in AI-assisted decision-making when AI was consistently reliable, with effects varying by explanation completeness and decision-making phase [1]. Specifically, in deliberative-phase evaluation, any level of reconstructive explanation reduced the need for critical engagement compared to no AI support by partially substituting for professionals' own evaluative processes. However, during decision-execution, minimal reconstructive explanations (partial post-hoc rationales requiring professionals to draw on internal schemas to complete AI's reasoning) sustained deeper analytical engagement and greater experienced decision-confidence, in contrast to fully reconstructive explanations which fostered a sense of decision-confidence dissociated from the analytical effort that would otherwise ground it [1].

The present work-in-progress extends these findings by introducing a *panoptic condition*, reflecting institutional oversight practices enforced by statutory bodies that subject professional judgment to the permanent possibility of scrutiny [5], and investigates whether these dynamics of critical thinking established across no-AI support and varying levels of reconstructive causal explainability completeness are sustained. We therefore ask: *How do the psychophysiological and behavioral dynamics of critical thinking manifest under panoptic conditions, and to what extent do they differ from those established under non-panoptic conditions?*

To address this question, we replicate the NeuroIS experimental methods of [1] under this panoptic condition. The preliminary findings, based on the analysis of electrodermal activity (EDA), behavioral, and self-report measures of N = 14 participants, suggest that panoptic conditions may preserve deliberative-phase but alter executive-phase dynamics of critical thinking.

Theoretical Background

Central to the exercise of independent reflective professional judgment that anchors responsibility and accountability in deontologically-governed professions, critical thinking is a self-directed, self-correcting, criteria-driven, and context-sensitive mode of thought conducive to informed judgment and calibrated decision-confidence under epistemic uncertainty [6–9], operating dynamically across the deliberative and executive phases of decision-making [10].

Using a multi-method NeuroIS approach [11] with practicing human resources (HR) professionals tasked with AI-assisted candidate pre-selection under consistently reliable AI, we showed that the level of completeness of reconstructive explanations influences the dynamics of critical thinking across decision-making phases in nuanced ways [1].

During the deliberative phase, minimal and fully reconstructive explanations reduced neural markers of sustained and selective attention compared to no-AI support, while reading times decreased with explanation completeness, consistent with more efficient predictive encoding and suggesting that any level of reconstructive explanations partially substitutes for internal evaluative processes, diminishing the need for critical thinking during the deliberative phase. However, during the executive phase, where judgment formation requires professionals to draw on internal schemas to complete AI's reasoning process, minimal reconstructive explanations were associated with greater activation of neural correlates of analytical reasoning and greater experienced decision-confidence. Decision times were longest with minimal reconstructive explanations, suggesting deeper analytical engagement consistent with systematic processing. Self-reported decision-confidence, by contrast, followed explanation completeness, dissociating from experienced decision-confidence, suggesting that minimal reconstructive explanations may better support the reasoning processes that ground experienced decision-confidence in analytical rigor. No significant differences in selective reliance were found between explanation conditions. Therefore, when AI outputs align with deontological standards, minimal reconstructive explanations scaffold, rather than substitute, internal reasoning, helping retain control over judgment and align confidence with analytical effort.

However, these dynamics unfold under the permanent possibility of institutional scrutiny. Deontologically-governed professions commonly rely on institutionalized oversight practices mandated by statutory bodies, such as audits, to create formal accountability and responsibility for standards-aligned professional judgment [12].

These institutionalized oversight practices, while varying in form across regulatory traditions, share the stated objective of protecting public interest through structured normative authority [13]. In doing so, such practices instantiate a *panoptic mechanism*, whereby the permanent possibility of audit sustains an organized field of visibility in which professionals know their conduct may be scrutinized at any time, inducing the internalization of norms and self-disciplining conformity [5].

Prior accountability research suggests that the permanent possibility of scrutiny whose specific focus cannot be anticipated may prompt preemptive self-criticism and more effortful, integrative processing, as individuals prepare to defend their judgment against diverse possible objections [14, 15]. Applied to AI-assisted professional decision-making, however, this effortful processing may follow divergent paths. Panopticism may *extend preemptive self-criticism to AI's reconstructive explanations*, as professionals anticipate having to justify the rationale underlying their decisions, therefore, amplifying the scaffolding of internal reasoning observed with minimal reconstructive explanations under non-panoptic conditions [1]. Alternatively, panopticism may impose a *compound cognitive demand*, whereby professionals must simultaneously evaluate the evidence at hand and AI's reconstructive explanation while managing awareness that their judgment may be subject to institutional scrutiny, potentially exceeding the cognitive resources that sustained the dynamics of critical thinking established under non-panoptic conditions [1]. We therefore posit that the panoptic mechanism may either amplify critical engagement, or impose a compound cognitive demand that differentially affects the no-AI support condition and AI-assisted reconstructive explanation conditions.

Methods

Data collection is ongoing, with the aim of matching the sample size of Experiment 1 in our previous study (N = 23) [1]. The present analysis draws on 14 active HR professionals, registered members of a recognized professional order (aged 24–45; M = 33; SD = 6.34; 78.6% female; average 9 years of experience). Participants received CAD100 compensation and met eligibility criteria for electroencephalography (EEG) and EDA measurements. The experiment was conducted at a North American university NeuroIS laboratory and approved by the Research Ethics Board (certificate #2023–5041). All participants provided written informed consent.

Following the design of Experiment 1 in our previous work [1], using identical stimuli, participants completed a scenario-based Wizard-of-Oz [16] AI-assisted recruitment task with consistently reliable AI, in which they pre-selected Web Analyst candidates from pairs of resumes based on three criteria, across 30 counterbalanced trials divided equally into three within-subject conditions: no AI support (NASP), minimal reconstructive explanations (MREP), and fully reconstructive explanations (FREP). Departing from Experiment 1, participants were informed following task instructions and practice trials that their decisions would be recorded and that a random subset could be audited post-hoc by an independent committee of HR experts against the professional codes governing their practice, with justification potentially required, a manipulation of anticipated scrutiny shown to produce observable cognitive and neurophysiological effects [17–19]. Systematic differences on the dependent variables are therefore attributable to the panoptic manipulation, as protocol, stimuli, and recruitment criteria were otherwise held constant across our two experiments [20]. Upon completion of the study, participants were informed that no such audit would be conducted.

Measures, materials, apparatus, data preprocessing, and statistical analyses replicated Experiment 1 [1]. This work-in-progress reports skin conductance responses (SCR), as established biomarkers of experienced decision-confidence [21, 22], reading and decision times, selective reliance and self-reported decision-confidence, following established practice of integrating psychophysiological and behavioral measures in NeuroIS research [23, 24]. EEG data and between-experiment comparisons are deferred to the final analysis upon completion of data collection.

Results

The preliminary analysis reports that, during the deliberative phase, reading times decreased significantly across conditions (NASP > MREP, $t(139) = 3.67$, $p < 0.001$, $d = 0.31$; NASP

> FREP, $t(139) = 6.80$, $p < 0.001$, $d = 0.58$; MREP > FREP, $t(139) = 3.34$, $p < 0.001$, $d = 0.28$).

During the executive phase, SCR amplitudes were significantly lower in MREP than FREP ($Z = -2.658$, $p = 0.008$), with NASP showing the lowest amplitudes, though non-significant. Self-reported decision-confidence was significantly higher in FREP than both NASP ($t(139) = -4.38$, $p < 0.001$, $d = 0.37$) and MREP ($t(139) = -3.12$, $p = 0.001$, $d = 0.26$), while the difference between NASP and MREP approached significance ($t(139) = -1.63$, $p = 0.053$, $d = 0.14$). Decision times showed a non-significant curvilinear trend, with the shortest average in MREP and the longest in FREP. Selective reliance was significantly lower in NASP compared to both MREP and FREP ($p < 0.001$). However, the difference between MREP and FREP was not significant.

Discussion

Taken together, these preliminary results offer initial insights into how the psychophysiological and behavioral dynamics of critical thinking manifest across the phases of decision-making under panoptic conditions, and how they differ from those established under non-panoptic conditions [1].

During the deliberative phase (résumés assessment), reading times decreased significantly across conditions, suggesting that predictive encoding dynamics observed under non-panoptic conditions [1] may be maintained under panoptic conditions.

During the executive phase (candidate pre-selection), the dissociation between experienced and self-reported decision-confidence persists: SCR amplitudes were significantly lower with minimal than fully reconstructive explanations, while self-reported decision-confidence followed explanation completeness. Thereby suggesting that minimal reconstructive explanations may continue to sustain experienced decision-confidence grounded in analytical rigor. However, SCR amplitudes were lowest without AI support, though non-significant, indicating that experienced decision-confidence may be greatest when professionals reason independently under panoptic conditions, inverting the non-panoptic pattern where experienced confidence was greatest with minimal reconstructive explanations [1]. Panopticism may therefore most readily support experienced decision-confidence when professionals exercise the self-directed reasoning that institutional oversight was instituted to discipline [5]. Decision times showed a non-significant trend, with fully reconstructive explanations now producing the longest average, contrary to [1], a pattern that may be consistent with the expectation that institutional scrutiny amplifies engagement with the most complete justificatory account available [14, 15]. Selective reliance was significantly lower without AI support than under both explanation conditions, whereas no such differentiation was observed under non-panoptic conditions [1], pointing to institutional scrutiny differentiating AI-supported from independent reasoning rather than between levels of explanation completeness.

Together, these early findings, pending consolidation with the complete sample, suggest that panopticism may preserve deliberative-phase dynamics of critical thinking. However, during decision execution, panopticism may transform the scaffolding function of minimal reconstructive explanations established in [1] to be less effective. Instead, greater analytical engagement and experienced decision-confidence may be most evident when professionals reason independently.

Conclusion: Implications and Next Steps

Our findings contribute to NeuroIS research by: (1) providing preliminary psychophysiological and behavioral observations suggesting that panopticism may preserve the deliberative dynamics of critical thinking while transforming its executive dynamics, such that analytical engagement and experienced decision-confidence may be most evident when professionals reason independently of AI support; (2) responding to the call for strengthening ecological validity of laboratory findings [25] through the recruitment of registered professionals and by experimentally reproducing the institutional oversight conditions that characterize deontologically-governed practice; and (3) extending the investigation of the cognitive and affective mediating mechanisms of IS use [26] to AI-assisted decision-making in deontologically-governed professions.

For practice, these findings suggest that institutional oversight practices designed for self-directed professional judgment may not straightforwardly extend to AI-assisted decision-making.

Future work will complete data collection to match the target sample size of $N = 23$ as per Experiment 1 [1], analyze EEG data, and conduct between-experiment comparisons with sufficient power to discriminate between the alternative paths posited in this study.

References

1. Korosec-Serfaty, M., Léger, P.-M., Parent-Rocheleau, X., Sénécal, S.: Critical thinking in AI-assisted deontologically-governed professional decision-making: When and how explainability, reliability, and transparency matter. Information Systems Frontiers. (2026). https://doi.org/10.1007/s10796-025-10691-
2. Arendt, H.: Thinking and moral considerations: A lecture. Soc. Res. (New. York). 38, 417–446 (1971).
3. Benbya, H., Pachidi, S., Jarvenpaa, S.L.: Special issue editorial: Artificial intelligence in organizations: Implications for information systems research, (2021). https://doi.org/10.17705/1jais.00662.

4. Lebovitz, S., Lifshitz-Assaf, H., Levina, N.: 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 (2022). https://doi.org/10.1287/ORSC.2021.1549.
5. Foucault, M.: Discipline and punish: The birth of the prison. Second Vintage Books Editions (2012).
6. Lipman, M.: Critical thinking: What can it be? Analytic Teaching. 8, (1987).
7. Cohen, M.S., Freeman, J.T.: Thinking naturally about uncertainty. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 179–183. Sage CA: Los Angeles, CA: SAGE Publications. (1996).
8. Siegel, H.: Educating reason: Rationality, critical thinking and education. Routledge Publishing, New York (1989).
9. Facione, P.A.: Critical thinking: What it is and why it counts. Insight Assessment. (2015). https://doi.org/https://doi.org/ISBN13:978-1-891557-07-1.
10. Facione, P.A.: Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research Findings and Recommendations. (1990).
11. Kirwan, C.B., Vance, A., Jenkins, J.L., Anderson, B.B.: Embracing brain and behaviour: Designing programs of complementary neurophysiological and behavioural studies. Information Systems Journal. 33, 324–349 (2023). https://doi.org/10.1111/isj.12402.
12. Freidson, E.: Professionalism: The third logic. University of Chicago Press, Chicago (2001).
13. Adams, T.L.: Regulating professions: The Emergence of professional self-regulation in four canadian provinces. University of Toronto Press (2018).
14. Lerner, J.S., Tetlock, P.E.: Accounting for the effects of accountability. Psychol. Bull. 125, 255–275 (1999).
15. Tetlock, P.E.: Accountability and complexity of thought. J. Pers. Soc. Psychol. 45, 74–83 (1983).
16. Riek, L.D.: Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. J. Hum. Robot. Interact. 1, 119–136 (2012). https://doi.org/10.5898/jhri.1.1.riek.
17. Tetlock, P.E., Kim, J.D.: Accountability and judgment processes in a personality prediction task. J. Pers. Soc. Psychol. 52, 700–709 (1987).
18. Balconi, M., Crivelli, D., Castiglioni, C., Lozza, E.: Social frame and tax compliance modulate electrophysiological and autonomic responses following tax-related decisions. Sci. Rep. 9, 4713 (2019). https://doi.org/10.1038/s41598-019-41156-7.
19. Hajcak, G., Moser, J.S., Yeung, N., Simons, R.F.: On the ERN and the significance of errors. Psychophysiology. 42, 151–160 (2005). https://doi.org/10.1111/j.1469-8986.2005.00270.x.
20. Shadish, W.R., Cook, T.D., Campbell, D.T.: Experimental and quasi-experimental designs for generalized causal Inference. Houghton Mifflin, Boston, MA (2002).
21. Bechara, A., Damasio, H., Tranel, D., Damasio, A.R.: Deciding advantageously before knowing the advantageous strategy. Science (1979). 275, 1293–1295 (1997).
22. Dawson, M.E., Schell, A.M., Courtney, C.G.: The skin conductance response, anticipation, and decision-making. J. Neurosci. Psychol. Econ. 4, 111–116 (2011). https://doi.org/10.1037/a0022619.
23. Riedl, R., Maedche, A.: Why behavioral measures as indirect measures of neurophysiological activity are important for progress in the NeuroIS discipline. In: Davis, F.D., Riedl, R., von Brocke, J., Léger, P.-M., and Müller-Putz, G.R. (eds.) Information systems and neuroscience:

NeuroIS 2025 (Lecture Notes in Information Systems and Organisation, Vol. 9). Springer (2025). https://doi.org/https://doi.org/10.1007/978-3-032-00815-2_15.

24. Dimoka, A., Davis, F.D., Gupta, A., Pavlou, P.A., Banker, R.D., Dennis, A.R., Ischebeck, A., Müller-Putz, G., Benbasat, I., Gefen, D., Kenning, P.H., Riedl, R., Vom Brocke, J., Weber, B.: On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS Quarterly. 36, 679–702 (2012).
25. Balapour, A., Riedl, R.: Ecological validity in NeuroIS research: Theory, evidence, and a roadmap for future studies. J. Assoc. Inf. Syst. 26, 9–65 (2025). https://doi.org/10.17705/1jais.00904.
26. vom Brocke, J., Hevner, A., Léger, P.M., Walla, P., Riedl, R.: Advancing a NeuroIS research agenda with four areas of societal contributions. European Journal of Information Systems. 29, 9–24 (2020). https://doi.org/10.1080/0960085X.2019.1708218.

Empathic Displacement in Algorithmic Decision-Making: A NeuroIS Study

Thomas Ware III[1], John-Patrick Akinyemi[2] and Lam An[3]

[1]University of Pittsburgh, Katz Graduate College of Business, Pittsburgh, PA
{thw55@pitt.edu}
[2]University of Texas-Austin, McCombs School of Business, Austin, Texas
{JohnPatrick.Akinyemi@mccombs.utexas.edu}
[3]University of Winnipeg, Department of Business and Administration, Winnipeg, Canada
{l.an@winnipeg.ca}

Abstract. AI recommendations increasingly anchor high-stakes social decisions, yet the neurological consequences for decision-makers' perspective-taking engagement during human-AI collaboration remain unexamined. We propose that AI anchoring operates through the attenuation of prefrontal cortex activity, the neural substrate of perspective-taking and deliberative reasoning, consequential when individuals are affected by the decision at hand. Using EEG and ERP paradigms in an undergraduate admissions context, this study aims to identify this neural mechanism and examines dispositional perspective-taking as a neurological buffer against AI-induced displacement.
Keywords: Algorithmic Decision-Making. Perspective-Taking. Prefrontal Cortex. Relational Perspective. NeuroIS.

Introduction

According to social psychology, when humans make consequential decisions, such as in admissions, hiring, lending, or criminal justice, they are embedded in a relational structure that extends beyond the mechanics of the decision itself [2]. At its center is the affected party: a person with goals, circumstances, and stakes whose future the decision will determine [2]. Resultantly, social cognitive theory suggests attention to the individual and her circumstances is salient when making decisions that will affect them [2]. AI is increasingly inserted into this structure through *algorithmic decision-making* (ADM) [12, 18], defined as decision processes where AI-generated outputs, such as recommendations, scores, or predictions, mediate the encounter between decision-makers and the individuals their judgments affect [13, 18].

Recent work conceptualizes AI as an external cognitive system (System 3) that can substitute for, or bypass, internal cognition through processes of *cognitive surrender* [15]. From this perspective, AI alters when and how decision-makers engage their own cognitive and social reasoning. Additional evidence suggests this phenomenon is not contingent on specific AI architecture. For instance, the ceding of evaluative judgment to AI-generated outputs is found across classical machine learning [14] and generative AI settings [10]. Thus, when decisions carry direct human consequences, understanding the effects of System 3 becomes critical [12]. If AI-induced cognitive surrender shifts engagement away from the affected party, individuals may be evaluated through algorithmic outputs rather than sustained consideration of their circumstances, leading to a de-humanized process [9]. In contexts where decisions carry direct consequences for individuals, we posit cognitive surrender carries an additional consequence; the displacement of the decision-maker's empathic orientation toward the affected party. Therefore ADM contexts with social consequence, where the decision-maker's orientation toward that subject is a meaningful component of the decision process [12, 18], is the concentration of our study.

We leverage an undergraduate admissions context for understanding how AI influences empathy in ADM. We operationalize ADM by employing a GenAI-augmented machine learning system that generates predictive recommendations. We furthermore leverage multimodal signal triangulation allows us to measure these effects [17] by synchronizing electroencephalography (EEG), eye-tracking, facial expression analysis, and galvanic skin response (GSR) to strengthen inferences about attenuated engagement during ADM. This method, along with time-locked ERP analysis, examines the temporal dynamics of neural activity relative to stimulus onset, allowing us to identify early windows of engagement that reflect automatic empathic orientation toward the individual [4, 11]. Importantly, this allows us to pinpoint *when* that change emerges, in the first moments of exposure, when an initial social-cognitive orientation takes shape, or later as deliberation unfolds.

Furthermore, we expect such effects may not be uniform across decision-makers [19]. Individuals with stronger dispositional perspective-taking (PT) may maintain engagement with the applicant despite AI's presence, while those with weaker empathic orientation may more readily defer to algorithmic outputs. Accordingly, we expect PT to reflect a stable orientation toward holding others' circumstances in mind [19], possessing a relational orientation that competes with, and may resist, the pull of the algorithm. These considerations motivate our research question:

Does AI support attenuate decision-makers' cognitive and empathic engagement with the applicant, and how does personality dispositions moderate this effect?

Theoretical Background

Empathic Engagement during ADM

In decision contexts with direct social consequences, the capacity to model another person's circumstances, motivations, and likely outcomes is the basis of sound judgment [2, 18]. Yet Shaw & Nave [15] demonstrate that when AI systems are used, decision-makers often engage in cognitive surrender, referred to as the uncritical adoption of AI outputs with minimal scrutiny. Effectively bypassing the deliberative and social-cognitive processes that sound judgment requires. This behavioral pattern reflects a reverse of agency in decision-making, from the person making the judgment to the algorithm [1]. When agency moves to the system, the applicant may also recede as a *person* in the decision-maker's experience, becoming less a lived narrative and more a set of model-relevant features. That matters because empathic engagement is not just "extra reasoning." It is a social orientation that shapes what information feels salient, what tradeoffs feel acceptable, and how much moral weight is afforded to the consequences borne by the affected party [6, 19]. Thus, AI-guided decision frames may dampen the *felt* connection and responsibility that typically motivate careful, human-centered evaluation.

Multimodal Measurement: Empathic Engagement and Temporal Dynamics

Empathic engagement with affected parties during consequential decisions draws on multiple cognitive and affective processes, including attention to socially meaningful information [6] and affective responsiveness [19]. However, we suggest when decision-makers cede judgment to algorithmic outputs, they disengage from the epistemic orientation that makes empathic appraisal possible. In turn, the affected party shifts from subjects whose circumstances demand interpretive engagement to objects already processed by the algorithm. We term this *empathic displacement*, defined by the systematic withdrawal of PT and contextual sensitivity through which cognitive surrender extends beyond decisional conformity to include moral distancing from the human consequences of the decision.

We study this effect as convergent reductions across four synchronized signals. While EEG captures the temporal signature of empathic attenuation [4, 11], it cannot independently confirm its source. Thus, we leverage additional measures to triangulate empathic displacement across attentional, affective and physiological measures. For instance, eye-tracking dwell time on applicant narrative content reveals whether attentional allocation shifts away from socially meaningful information toward algorithmic output during the early processing window [3]. Facial expression emotional responsiveness indicates whether affective engagement with the applicant persists or flattens under ADM exposure [7]. Additionally, GSR arousal during profile review provides a physiological measure of engagement independent of conscious report [16]. Together, this triangulation isolates empathic displacement as the operative mechanism rather than general disengagement or reduced cognitive effort.

Furthermore, the temporal dimension of neural engagement is particularly informative [11]. Time-locked ERP analysis allows us to examine when engagement shifts relative to stimulus onset. Early neural activity following exposure to an applicant profile reflects the initial orienting response to socially meaningful information, the rapid, largely automatic assessment that a person with circumstances and stakes is present [4, 11]. Adversely, later activity reflects deliberative evaluation, emotional regulation, and decision formation [4, 11]. Therefore, if ADM attenuates the early orienting response while leaving later activity relatively intact, this pattern is consistent with disrupted empathic engagement rather than general cognitive effort reduction. If attenuation is uniform across the full window, alternative explanations such as reduced effort or task disengagement become more plausible.

Hypothesis Development

Cognitive Surrender Orients Empathic Displacement

Prior behavioral work establishes that AI recommendations anchor decision outcomes [14], and Shaw & Nave [15] demonstrate that this anchoring often manifests as cognitive surrender. Yet cognitive surrender has been treated as a purely cognitive effect. However, we suggest when decision-makers cede judgment in ADM, they disengage from the epistemic orientation that makes empathic appraisal possible. In turn, the affected party shifts from a person whose story invites interpretation to an object treated as already decided by the algorithm. [9]. Resulting in an interpersonal distancing from the human consequences of the decision. Leading us to our first hypothesis;

H1: ADM will produce convergent attenuation in empathic engagement across multimodal signals (EEG, eye-tracking, facial expression, GSR) relative to non-AI advisory conditions.

Empathic Displacement Localizes to Early Neural Activity

Differences in empathic engagement should manifest in early-phase neural responses to socially meaningful stimuli, as Li & Han [11] demonstrated that early neural activity is selectively sensitive to socially relevant processing before later regulatory mechanisms engage. We leverage this temporal specificity to identify empathic displacement, expecting it to be evident in the attenuation of these early engagement signatures, reflecting a failure to orient toward the applicant's circumstances rather than general cognitive load reduction [4]. Triangulation across modalities strengthens this inference [3, 7, 16, 17], as convergent attenuation across neural, attentional, affective, and physiological channels would affirm that the observed pattern reflects systemic empathic disengagement rather than an artifact of any single measure. If empathic displacement operates at the level of orientation, its temporal signature should be similarly specific. Therefore, we hypothesize the following;

H2: ADM will produce greater attenuation in early-phase neural engagement with socially relevant stimuli than in later-phase regulatory activity

Dispositional PT Is a Buffer Against Empathic Displacement

If the previous two hypotheses hold, they would together indicate the presence of empathic displacement, yet such effects. Leveraging prior literature on empathy [19], we posit that those with strong empathic engagement capabilities attenuate the effects of cognitive surrender across decision outcomes. Thus, those with higher dispositional PT, as measured by the IRI, may serve as a buffer against empathic displacement. Not because high-IRI individuals are less susceptible to cognitive surrender behaviorally, but because their engagement with the affected party as a subject with a perspective is more resistant to displacement. We therefore predict that decision-makers with higher IRI scores will show correspondingly smaller reductions in early-phase neural engagement with socially relevant stimuli under ADM. Thus, our final hypothesis is stated as follows;

H3: Higher dispositional PT (IRI) will moderate empathic displacement in ADM

Experiment Design

Participants

We aim to recruit 25–35 graduate students (master's, law, PhD, and Ed.D candidates), with an initial trial of approximately 10 participants. This group is theoretically meaningful for testing empathic displacement. Specifically, p articipants are equipped with domain familiarity and intrinsic investment in educational outcomes needed to establish baseline engagement with the applicant [8]. This allows us to focus on participants who would be expected to attend to the affected party absent AI, allowing any displacement to be more plausibly attributed to ADM.

Experimental Context: Undergraduate College Admissions

We select undergraduate admissions for our research context, as decisions are consequential for applicants' life trajectories, where ADM is becoming increasingly involved [12, 18]. Thus, making this setting ecologically appropriate for studying empathic displacement. Applicant profiles are designed to reliably elicit empathy across conditions. Thus, each profile must include a substantive contextual background that presents the applicant as a person with goals, circumstances, and potential.

Pre-Experimental Survey

Prior to the experiment, participants will complete the PT sub-scale of the Interpersonal Reactivity Index (IRI) [5], which serves as the primary individual difference moderator in H3 analyses. The subscale is theoretically appropriate because it measures dispositional PT [5], referring to an individual's stable, trait-level tendency to adopt another person's point of view. This represents a natural inclination rather than a situationally induced state using self-reported measure. Therefore, by mapping IRI scores onto empathic engagement during the decision encounter, we can examine whether those who naturally tend toward PT show greater neural resistance to empathic displacement.

Procedure

Participants are randomly assigned to one of two between-subjects conditions:

- **Non-AI condition**: Profiles are displayed alongside a *Success Score*, rule-based algorithmic-driven score
- **AI condition**: Profiles are displayed alongside an *AI recommendation* generated by a GenAI-augmented, machine learning model trained on application data.

The non-AI condition provides equivalent informational input through rule-based score that carries no evaluative authority over the decision. The GenAI-augmented ML output operates differently. Trained on applicant data, it derives its output from learned patterns and presents as a judgment that implies the applicant has already been assessed. By holding informational content constant, while varying the authority of the advisory input, this design allows us to attribute differences in empathic engagement to the AI condition rather than informational asymmetry. Accordingly, each trial follows the sequence: fixation cross (500ms) → applicant profile with advisory input (3 minutes) → decision prompt (Admit / Reject / Waitlist, self-paced) → inter-trial interval (1–2 seconds), repeated 20 times per participant with continuous EEG recording. Post-task questionnaires will capture decision experience, deliberation patterns, and subjective engagement with applicant profiles.

Data Analysis

Data analysis will employ a triangulation approach using the iMotions multimodal research platform, which synchronizes time-locked data collection across EEG, eye-tracking, facial expression analysis, and GSR within a unified temporal framework. This synchronized multimodal design enables convergent analysis across neural engagement markers, attentional allocation, affective responsiveness, and autonomic arousal (H1), examination of early- versus late-phase temporal dynamics in neural engagement (H2), and moderation analysis of dispositional perspective-taking across these convergent indicators (H3).

Discussion

This study makes two distinct contributions to the literature on ADM. The first is the identification and characterization of the mechanism through which ADM reorganizes decision-making in contexts where decisions carry direct human consequence, require judgment under uncertainty, and resist full algorithmic resolution [12, 18]. Prior work has established that AI functions as a System 3 cognitive anchor [14], manifests as cognitive surrender [15]. We extend this line of research by demonstrating that in ADM contexts with direct social consequences, cognitive surrender is not merely a failure in deliberation, but displaces the empathic engagement that connects decision-makers to the affected party. We identify such effects through a triangulation across neural, attentional, affective, and physiological signals. Resultantly, allowing us to be more precise about connecting the behavioral phenomenon of cognitive surrender to its empathic mechanism.

The second contribution is the establishment of dispositional PT as a buffer against empathic displacement via IRI moderation analysis [5]. These contributions together address the normative motivation underlying this line of research: re-humanizing a decision process that ADM systematically de-humanizes. In this view, PT acts as a salient sensemaking capacity that enables decision-makers to hold the affected party's circumstances, context, and subjectivity in view, contributing to ADM processes in ways that algorithmic outputs cannot [12, 18]. Understanding the mechanism of its displacement, and identifying who is most and least susceptible to that displacement, is therefore a necessary precondition for designing interventions that preserve the distinctly human capacities that high-stakes social decisions demand.

References

1. Angelopoulos, S., Bendoly, E., Fransoo, J., Hoberg, K., Ou, C., Tenhiälä, A.: Digital transformation in operations management: Fundamental change through agency reversal. Journal of Operations Management 69, 876–889 (2023). https://doi.org/10.1002/joom.1271
2. Bandura, A.: Social cognitive theory: An agentic perspective. Annual Review of Psychology 52(1), 1–26 (2001). https://doi.org/10.1146/annurev.psych.52.1.1
3. Bera, P., Soffer, P., Parsons, J.: Using eye tracking to expose cognitive processes in understanding conceptual models. MIS Quarterly 43(4), 1105–1126 (2019). https://doi.org/10.25300/MISQ/2019/14163
4. Cheng, Y., Chen, C., Decety, J.: An EEG/ERP investigation of the development of empathy in early and middle childhood. Developmental Cognitive Neuroscience 10, 160–169 (2014). https://doi.org/10.1016/j.dcn.2014.08.012
5. Davis, M.H.: Measuring individual differences in empathy: Evidence for a multidimensional approach. Journal of Personality and Social Psychology 44(1), 113–126 (1983). https://doi.org/10.1037/0022-3514.44.1.113
6. Decety, J., Lamm, C.: Human empathy through the lens of social neuroscience. The Scientific World Journal 6, 1146–1163 (2006). https://doi.org/10.1100/tsw.2006.221
7. Dimberg, U., Andréasson, P., Thunberg, M.: Emotional empathy and facial reactions to facial expressions. Journal of Psychophysiology 25(1), 26–31 (2011). https://doi.org/10.1027/0269-8803/a000029
8. Hegarty, N.: Application of the Academic Motivation Scale to graduate school students. The Journal of Human Resource and Adult Learning 7(2), 73–81 (2011)
9. Kim, H., McGill, A.: AI-induced dehumanization. Journal of Consumer Psychology 35, 363–381 (2024). https://doi.org/10.1002/jcpy.1441
10. Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., Wilson, N.: The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), Article 1121, pp. 1–22. ACM, New York (2025). https://doi.org/10.1145/3706598.3713778
11. Li, W., Han, S.: Perspective taking modulates event-related potentials to perceived pain. Neuroscience Letters 469(3), 328–332 (2010). https://doi.org/10.1016/j.neulet.2009.12.021
12. Melville, N.P., Robert, L., Xiao, X.: Putting humans back in the loop: An affordance conceptualization of the 4th industrial revolution. Information Systems Journal 33(4), 733–757 (2023). https://doi.org/10.1111/isj.12422

13. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: An attentional integration. Human Factors 52(3), 381–410 (2010). https://doi.org/10.1177/0018720810376055
14. Rhue, L.: The anchoring effect, algorithmic fairness, and the limits of information transparency for emotion artificial intelligence. Information Systems Research 35(3), 1479–1496 (2023). https://doi.org/10.1287/isre.2019.0493
15. Shaw, S.D., Nave, G.: Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender. The Wharton School Research Paper (2026). https://doi.org/10.2139/ssrn.6097646
16. Shi, Y., Ruiz, N., Taib, R., Choi, E., Chen, F.: Galvanic skin response (GSR) as an index of cognitive load. In: CHI '07 Extended Abstracts on Human Factors in Computing Systems, pp. 2651–2656. ACM, New York (2007). https://doi.org/10.1145/1240866.1241057
17. Tams, S., Hill, K., Ortiz de Guinea, A., Thatcher, J., Grover, V.: 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(10), 723–753 (2014). https://doi.org/10.17705/1jais.00374
18. Teodorescu, M.H.M., Morse, L., Awwad, Y., Kane, G.C.: Failures of fairness in automation require a deeper understanding of human–ML augmentation. MIS Quarterly 46(2), 1287–1304 (2022). https://doi.org/10.25300/MISQ/2021/16535
19. Van Boven, L., Loewenstein, G.: Empathy gaps in emotional perspective taking. Current Directions in Psychological Science 14(2), 79–83 (2005). https://doi.org/10.1111/j.0963-7214.2005.00340.x

Neural Uncertainty Monitoring During Repeated GenAI Interaction: A Pilot EEG Study Using Frontal Theta Power

Thomas Kalischko[1], Stefan Eibl[1], Michael Hollaus[1], René Riedl[2,3]

[1] University of Applied Sciences Wiener Neustadt, Wieselburg, Austria
(thomas.kalischko,stefan.eibl,michael.hollaus)@fhwn.ac.at
[2] University of Applied Sciences Upper Austria, Steyr, Austria
rene.riedl@fh-steyr.at
[3] Johannes Kepler University Linz, Linz, Austria

Abstract. This pilot study examines whether neural uncertainty monitoring changes across repeated interactions with Generative Artificial Intelligence (GenAI) and whether a final-stage explanation modulates this process. Using EEG time-frequency analysis, we quantified frontal theta power (4-7 Hz) at F3/F4 across three interaction moments: the first AI response (M1), the second AI response (M2), and the final decision-relevant moment (Final). Frontal theta decreased significantly from M1 to Final, consistent with reduced uncertainty monitoring across repeated GenAI interaction. Intermediate contrasts were not significant, suggesting that neural adaptation may be concentrated at the final decision-relevant stage rather than unfolding linearly. Exploratory analyses of the Final moment indicated a context-dependent tendency in the effect of explanation, but this pattern did not reach the corrected significance threshold and should therefore be interpreted cautiously. Overall, the findings provide preliminary NeuroIS evidence that repeated GenAI interaction may be accompanied by reduced frontal uncertainty monitoring.

Keywords: Electroencephalography (EEG), Trust, NeuroIS, Generative AI (GenAI), Uncertainty Monitoring, Brain, Frontal Theta

Introduction

The rapid advancement of Generative Artificial Intelligence (GenAI) is fundamentally transforming the digital commerce landscape, shifting the focus from simple search algorithms to sophisticated virtual advisors that generate highly personalized content [1–3]. In high-involvement domains such as booking hotels or low-involvement domains such as purchasing consumer electronics, GenAI systems act as agentic entities that analyze complex user needs to provide realistic recommendations, effectively bridging the "knowledge and intelligence gap" between automated systems and human reasoning [4]. While these capabilities offer unprecedented efficiency in reducing information overload, they also introduce significant psychological challenges, as users must navigate the risks of hallucinations and potential biases in generated product suggestions [3, 4]. Trust is central for adopting GenAI in high-involvement shopping, and trust and distrust should be treated as distinct constructs that can co-occur [2, 5, 6]. Empirical evidence from adjacent human–technology contexts shows that perceived opacity in digital systems can simultaneously erode trust, heighten privacy concerns, and increase stress, dynamics that may translate analogously to consumer interactions with opaque GenAI systems [7]. Although prior research has established the importance of trust in AI and automation, trust is too often measured by a small set of popular self-report measures, and self-report generally requires pausing the task, meaning surveys are often infrequently administered and therefore may not capture the full evolution of trust [8]. Self-reporting methods, such as questionnaires measuring trust, capture static aspects of trust but cannot reflect its dynamic nature [9]. As a result, less is known about how trust-related monitoring processes evolve during repeated interaction with GenAI in real time. As trust is a dynamic and task-dependent concept, new methods are required to infer or predict a person's trust in an agent over time, given their interactions, rather than using post-hoc questionnaires to elicit trust [10]. This limitation is particularly relevant for GenAI systems, whose outputs can vary across prompts and even across repeated runs, making consistency and predictability less stable than in many traditional automation settings. The generative process is largely opaque to users, who cannot inspect the underlying reasoning and therefore lean on surface cues such as fluency, coherence, and conversational presentation when judging outputs, cues that can inflate perceived credibility and reliance even when responses are partially inaccurate [11]. Their responses can change with updates, the details of which are often not made transparent by LLM providers, and outputs can be nondeterministic in the sense that the same prompt leads to a different response when input to the model again, making the behavior of such models difficult to predict [12]. These conversational and justificatory responses may thus dynamically shape users' perceptions of predictability, transparency, and reliance. To address these limitations, the interdisciplinary field of Neuro-Information Systems (NeuroIS) employs neurophysiological tools and knowledge to gain more objective insights into the cognitive and emotional processes underlying trust [13, 14]. Prior NeuroIS research has predominantly relied on retrospective, survey-based approaches rather than continuous neurophysiological measurement [15]. Among these tools, Electroencephalography (EEG) has emerged as a major instrument for investigating the neural mechanisms of cognitive load, decision-making biases, and "cognitive offloading" which defines the process by which individuals delegate mental effort to GenAI systems during complex tasks like comparing hotel options or evaluating consumer electronics [13].

Theory and Hypothesis Development

Trust in GenAI as a Dynamic Calibration Process

Trust in AI is not established instantaneously but develops through repeated interaction and ongoing calibration. In trust theory, trust refers to a willingness to accept vulnerability based on positive expectations regarding another party's competence and intentions [16]. Applied to technological systems, this willingness is not fixed but is continuously updated as users accumulate experience [17]. In human–AI interaction, such expectations are formed and updated as users observe the system's behavior over time [18, 19]. This dynamic process has been characterized as trust calibration — the ongoing alignment between subjective trust and the actual reliability of an automated system [18, 20]. This process is especially relevant for Generative AI (GenAI), because GenAI systems do not merely retrieve predefined options but generate conversational, context-sensitive outputs that may appear reasoned, adaptive, and responsive [13, 21]. Unlike deterministic software, GenAI systems can produce variable and non-transparent responses, which increases the challenge of forming stable expectations and relying on prior experience to guide trust [22]. Across repeated interaction, users gradually accumulate cues about the system's predictability, usefulness, and apparent competence. According to Information Richness Theory, equivocality declines when additional cues reduce ambiguity and support interpretation [23]. Similarly, research on trust in automation emphasizes that trust is calibrated dynamically as users gain experience with system performance [18, 19]. Applied to GenAI, this suggests that uncertainty may be highest during the initial interaction, when users are still evaluating the quality and reliability of the system's output, and lower at later stages, once repeated responses provide a basis for expectation formation. This pattern mirrors findings in the broader human–AI interaction literature, which suggest that initial uncertainty gradually decreases as users acquire diagnostic information about system behavior [24].

Frontal Theta as a Neural Marker of Uncertainty Monitoring

To capture this dynamic calibration process at the neural level, we focus on frontal theta power. Frontal theta activity in the 4–7 Hz range has repeatedly been linked to uncertainty monitoring, conflict processing, and cognitive control [25–27]. It is particularly relevant when individuals evaluate ambiguous information, detect conflict, or allocate monitoring resources to guide action selection [25, 26]. Frontal midline theta is understood to reflect a domain-general computation of the need for control — sensitive to novelty, uncertainty, conflict, and error across contexts [25]. In the context of GenAI interaction, early responses from the system may evoke stronger uncertainty monitoring because users have not yet established stable expectations regarding output quality or reliability. As interaction continues and the system becomes more predictable, monitoring demands may decline, consistent with evidence that frontal theta decreases as stimulus patterns become familiar and expectations stabilize [28]. Accordingly, frontal theta provides a theoretically grounded neural indicator for examining whether repeated GenAI interaction is associated with reduced uncertainty monitoring. Based on this reasoning, we propose the following hypothesis:

H1: Frontal theta power is lower at the final decision-relevant moment (Final) than at the first AI response (M1).

Repeated Interaction Trajectory

Although a lower level of uncertainty monitoring is expected at the final moment than at the beginning of the interaction, the exact trajectory across intermediate moments is less clear. Two competing accounts can be distinguished. One possibility is a gradual, monotonic decline, such that each additional GenAI response incrementally reduces ambiguity, consistent with associative learning accounts that predict a progressive attenuation of prediction error signals as stimuli and their consequences become more familiar [29]. From a neural standpoint, this account is supported by evidence that frontal midline theta power systematically decreases as the brain forms increasingly stable representations of recurring stimuli — a pattern observed across habituation paradigms and interpreted as reflecting reduced demands for cognitive monitoring once outcomes become predictable [25, 30]. A second possibility is a more concentrated shift, in which uncertainty remains relatively stable across early interaction moments and only substantially declines once the interaction reaches a decision-relevant stage — that is, once users have accumulated sufficient diagnostic evidence to anchor their expectations about the system's reliability [18, 31]. This account aligns with research showing that trust calibration in human–automation interaction does not always develop smoothly but can instead consolidate rapidly once users reach a threshold of system familiarity or encounter decision-critical output [19, 32]. The distinction between these trajectories matters theoretically: a monotonic decline would suggest that each GenAI response carries approximately equal informational weight in reducing uncertainty, whereas a concentrated late-stage shift would imply that earlier responses serve primarily an exploratory function, exposing users to the system's style and apparent reasoning, while the final, decision-relevant response triggers the critical recalibration of monitoring demands. Because repeated exposure provides cumulative evidence about system behavior regardless of the exact trajectory, we nevertheless expect an overall decrease in frontal theta across the interaction sequence. Accordingly, we propose the following hypothesis:

H2: Frontal theta shows a decreasing trajectory across repeated interaction moments (M1, M2, Final).

Final-Stage Explanations and Context

Explanations are commonly assumed to reduce opacity by making system output more interpretable [33, 34]. In GenAI settings, explanations may also be perceived as relational signals because justificatory language can resemble interpersonal accountability and responsiveness [35–37]. This effect may be amplified by the conversational design of GenAI systems, which tends to elicit social processing and parasocial responses even when users are aware of the system's non-human nature [37, 38]. Such explanations may therefore reduce ambiguity and support reliance, especially when users face consequential or involvement-rich decisions. At the same time, explanation effects may depend on application context. In higher-involvement contexts, users may perceive greater vulnerability and may therefore be more likely to value additional justificatory cues [16, 18]. In lower-involvement contexts, the same cues may carry less diagnostic value because users may rely more strongly on heuristics and require less extensive evaluation, consistent with dual-process accounts of information processing [39]. Because explanations were introduced only at the Final moment in the present design, we treat their effect, and its possible dependence on context, as exploratory. Accordingly, we examine the following exploratory research question:

RQ1: Does the effect of a final-stage explanation on frontal theta depend on application context?

Methodology

Participants, Study Design, and Procedure

Ten participants (N = 10) took part in a laboratory EEG pilot study examining trust-relevant evaluation of GenAI recommendations across repeated interaction moments. The small sample reflects the resource-intensive nature of EEG data collection and the exploratory character of the study, which aimed to provide preliminary effect-size estimates rather than population-level conclusions. The experiment employed a mixed design with Context (hotel booking vs. smartphone search) as a between-subject factor and Explanation at Final (present vs. absent) as a within-subject factor. The two contexts reflected different consumer decision settings: a higher-involvement context (hotel booking) and a lower-involvement context (smartphone search). Each participant completed two counterbalanced stimulus blocks within their assigned context, differing only in whether the final GenAI output included an explanation. Block order was randomized. Participants interacted with a custom GPT presenting identical, pre-defined AI outputs within their condition, ensuring stimulus standardization. Upon arrival, participants were seated, fitted with an EEG cap, and completed a 2-minute resting baseline (60 s eyes open, 60 s eyes closed) before the task. Event markers were recorded throughout for later segmentation. Within each block, participants encountered three predefined interaction moments — M1, M2, and Final — representing successive AI responses, with explanations appearing exclusively at the Final moment. Participants were instructed to evaluate the outputs as realistic consumer decisions. After each block, participants completed a post-block questionnaire assessing social presence, trusting beliefs, perceived ease of use, perceived usefulness, usage intentions, and a manipulation check for perceived transparency/justification based on Qiu and Benbasat (2009)[40]. After both blocks, participants completed a final Trust-in-AI / Distrust-in-AI questionnaire capturing more general attitudes toward AI systems based on Scharowski et al. (2025)[41]. The session ended with debriefing.

Event Structure and Analytical Unit

The experiment yielded repeated EEG observations for the three focal interaction moments (M1, M2, Final) across two blocks per participant. Because each participant completed two blocks and each block contained one occurrence of each interaction moment, each participant contributed two observations per interaction moment at the block level. For EEG feature extraction, event-related spectral measures were first computed separately for each occurrence of M1, M2, and Final. These values were then averaged across the two occurrences of the same interaction moment within each participant. Consequently, inferential statistics were conducted at the participant level (N = 10) rather than at the level of single trials or epochs. This aggregation strategy was chosen to avoid pseudo-replication and to ensure that the statistical unit of analysis matched the experimental unit. EEG epochs were time-locked to AI-response onset, which served as the event of interest for the task-related analyses. For each participant, three response-locked epochs were obtained per block, corresponding to the first AI response (M1), the second AI response (M2), and the final decision-relevant AI response (Final). Because each participant completed two blocks, this yielded a theoretical maximum of six task-related epochs per participant, that is, two epochs for each interaction moment. Across the full sample of 10 participants, the theoretical maximum was therefore 60 task-related epochs (20 for M1, 20 for M2, and 20 for Final) before artifact-based epoch rejection. For inferential testing, epochs were averaged within participant and interaction moment, resulting in one aggregated value per participant for M1, M2, and Final.

EEG Recording and Preprocessing

EEG was recorded using a Brain Products X.on system with a sparse 7-channel montage consisting of F3, F4, C3, Cz, C4, P3, and P4. An auxiliary bipolar channel was recorded but not included in the present analyses. Data were processed in EEGLAB. Continuous EEG was filtered using a 0.5-30 Hz zero-phase FIR filter and re-referenced to the average reference. Independent Component Analysis (ICA) was then applied for artifact correction. Components reflecting ocular or muscular activity were identified based on their scalp topography, activation time course, and spectral characteristics and were removed prior to time-frequency analysis. EEG data were segmented relative to AI-output onset, which served as the primary event marker for the neural analyses. Epochs ranged from -200 to 1000 ms relative to AI-output onset. Baseline correction used the -200 to 0 ms prestimulus interval. The primary analysis window for feature extraction was the 200-800 ms post-stimulus interval, which was selected to capture post-stimulus uncertainty-monitoring activity while excluding the pre-stimulus baseline period. Importantly, this event-related analysis window refers to a short neural segment within the broader stimulus block, not to the duration of the entire block.

Time-Frequency Analysis and Neural Measures

Time-frequency analysis focused on event-related spectral perturbations (ERSPs) computed using EEGLAB's newtimef function in the 3-30 Hz range with baseline correction. The primary neural outcome was frontal theta power (4-7 Hz) averaged across electrodes F3 and F4 and quantified as mean ERSP amplitude in the 200-800 ms post-stimulus window. This focus on frontal theta at F3/F4 was theory-driven rather than data-driven. Frontal theta has been consistently linked to uncertainty monitoring, conflict processing, and cognitive control, which directly corresponds to the central theoretical construct of the present study [25–27]. Because theta power is maximal over frontal midline sensors, electrode sites in the frontal region — including F3, Fz, and F4 — have been selected a priori in prior EEG research based on this established topographic pattern [42]. This localization is further supported by evidence that mid-frontal theta activities reflect a common computation used for realizing the need for cognitive control, and that other frequency bands, while potentially informative, serve distinct roles that are separable from the core uncertainty and monitoring signal indexed by frontal theta [25]. Consistent with this, mid-frontal ERP components and theta band oscillatory processes reflect a common underlying action monitoring mechanism that is sensitive to novelty, conflict, punishment, and error across different task demands [43]. Prefrontal theta oscillations more broadly have been proposed to serve as a coordinating mechanism through which the prefrontal cortex recruits task-relevant networks during cognitively demanding situations, further underscoring their suitability as a theory-grounded marker [44]. Although EEG was recorded from seven scalp locations and additional frequency bands may also be informative, the pilot sample required a restricted primary analysis strategy in order to limit multiple testing and preserve interpretability. Restricting analysis to a pre-specified electrode cluster and frequency band is consistent with recommended practice in small-sample neuroimaging studies, where the number of statistical comparisons must be tightly controlled to maintain acceptable false positive rates [45]. Accordingly, the present manuscript treats frontal theta at F3/F4 as the confirmatory neural marker of interest.

Statistical Analysis

To test the primary hypothesis, paired-samples t-tests compared frontal theta between the key interaction moments M1 and Final. To examine the broader trajectory across repeated interaction, additional paired comparisons were conducted for M1 vs. M2 and M2 vs. Final. Because the study was based on a small pilot sample, corrected p-values were emphasized throughout and effect sizes were reported as paired-samples Cohen's d. To examine the exploratory role of explanation and context at the Final moment, a mixed analysis was conducted with Context (hotel vs. smartphone) as a between-subject factor and Explanation (present vs. absent) as a within-subject factor. Because explanations were manipulated only at the Final moment and the sample size was small, this analysis is explicitly treated as exploratory.

Self-Report Measures

Questionnaire data were analyzed descriptively. The post-block measures assessed social presence, trusting beliefs, perceived ease of use, perceived usefulness, usage intention, and perceived transparency/justification. The final questionnaire assessed general trust in AI and distrust in AI. For each item and construct, descriptive statistics were computed to document general response patterns and to verify whether the explanation manipulation increased perceived transparency/justification. No inferential statistics were reported for questionnaire data in the present pilot manuscript because the available export was aggregated at the item/condition level rather than stored in participant-level format suitable for inferential testing. Consequently, correlations between self-report measures and EEG measures were not computed in the present version and should be examined in future follow-up work using participant-level questionnaire data.

Results

Descriptive Overview of Frontal Theta Across Interaction Moments

Table 1 presents descriptive statistics for the primary neural outcome, frontal theta power (4–7 Hz) averaged across F3/F4, separately for the three interaction moments M1, M2, and Final. Descriptively, frontal theta was highest at the first AI response (M1) and lowest at the final decision-relevant moment (Final), indicating an overall decline across repeated GenAI interaction. The second interaction moment (M2) was located between these two points, suggesting a downward trend in the primary neural measure across the interaction sequence. Because each participant completed two blocks, each interaction moment was observed twice per participant. For inferential testing, the two AI-response-locked epochs per interaction moment were averaged within participant, resulting in one aggregated frontal-theta value per participant for M1, M2, and Final.

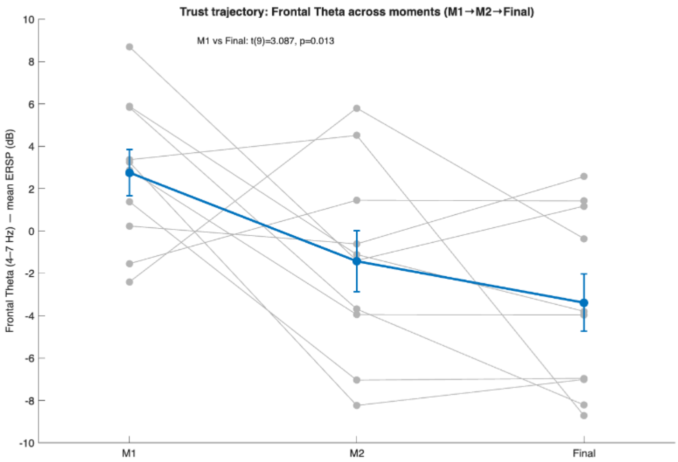


Fig. 1. Frontal Theta Across M1, M2, Final

Trust Dynamics Across Interaction Moments and the Role of Explanation

To test H1, paired-samples t-tests were used to compare frontal theta between the first AI response (M1) and the final decision-relevant moment (Final). The analysis revealed a significant reduction in frontal theta from M1 to Final, $t(9) = 3.09$, $p = .013$, corrected $p = .039$, Cohen's $d = 0.98$. Thus, frontal theta was lower at the final interaction moment than at the initial response, supporting H1. This result is consistent with the assumption that uncertainty-monitoring demands were lower at the end of the repeated GenAI interaction than at its beginning. Given the pilot nature of the study, this finding should nevertheless be interpreted as preliminary evidence.

To examine H2, additional paired comparisons were conducted across the full interaction sequence. The comparison between M1 and M2 did not reach statistical significance, $t(9) = 2.00$, $p = .077$, corrected $p = .153$, Cohen's $d = 0.63$. Likewise, the comparison between M2 and Final was not significant, $t(9) = 1.25$, $p = .243$, Cohen's $d = 0.40$. Taken together, these results do not provide full statistical support for a strictly decreasing moment-to-moment trajectory across M1, M2, and Final. While the descriptive pattern is consistent with an overall reduction in frontal theta across repeated interaction, only the contrast between the first and final moment reached significance. Therefore, H2 was not supported.

To address RQ1, we explored whether the effect of a final-stage explanation on frontal theta depended on application context. Because explanations were manipulated only at the Final moment and the sample size was small, this analysis was treated as exploratory. A mixed analysis with Context (hotel vs. smartphone) as a between-subject factor and Explanation (present vs. absent) as a within-subject factor was conducted for frontal theta at the Final moment. The Context × Explanation interaction did not reach the conventional significance threshold, $F(1,8) = 5.13$, $p = .053$. No robust main effects were observed. Accordingly, the data do not permit a confirmatory conclusion regarding context-dependent explanation effects. At most, the result indicates an exploratory numerical tendency that should be interpreted cautiously and examined in a larger follow-up study.

Self-Report Measures

The post-block questionnaire assessed social presence, trusting beliefs, perceived ease of use, perceived usefulness, usage intention, and perceived transparency/justification. In addition, participants completed a final Trust-in-AI / Distrust-in-AI questionnaire capturing more general attitudes toward AI systems. Descriptively, participants reported generally positive evaluations of the GenAI recommendations, particularly with regard to usefulness and advice quality. At the same time, general trust in AI and distrust in AI were both moderate (Trust: $M = 3.57$; Distrust: $M = 3.68$), suggesting a pattern of ambivalence rather than unconditional acceptance. The manipulation check further indicated higher perceived transparency/justification in the explanation condition than in the no-explanation condition, which supports the intended difference between the two Final-stage conditions. Because of the sample size, no inferential statistics or correlations between self-report and EEG variables were computed in the present pilot report. These associations should be examined in future work using participant-level questionnaire data.

Summary of Findings

Overall, the results provide preliminary evidence that frontal theta decreased from the first AI response to the final decision-relevant moment of repeated GenAI interaction. This finding supports H1. By contrast, the broader hypothesis of a strictly decreasing trajectory across all three interaction moments (H2) was not supported statistically, although the descriptive pattern was in the expected direction. Finally, the exploratory analysis for RQ1 did not yield conclusive evidence that the effect of a final-stage explanation depended on application context.

H	Description	Statistical Test	Result	Supported
H1	Frontal Theta is lower at Final than at M1	t(9)=3.09, p=.013, d=.98	Significant decrease	Yes
H2	Frontal Theta decreases across M1, M2, and Final	M1–M2: p=.077; M2–Final: p=.243	No full moment-to-moment decrease	No
H5	Does the effect of final-stage explanation depend on context?	ns F(1,8)=5.13, p=.053	Exploratory, inconclusive	No

Table 1. Summary of Hypotheses and Results.

Discussion and Conclusion

The present pilot study examined whether repeated interaction with GenAI is associated with changes in a neural marker of uncertainty monitoring. The main finding was a significant reduction in frontal theta from the first AI response to the final decision-relevant moment. This pattern is consistent with the idea that repeated GenAI interaction may reduce uncertainty monitoring as users accumulate cues about system behavior and calibrate reliance across the interaction sequence. Importantly, the data do not support a strong claim of a strictly monotonic decrease across all interaction moments. Instead, the reduction was most clearly visible between the first and final moment, suggesting that adaptation may become most pronounced once the interaction reaches a decision-relevant stage. This interpretation is compatible with trust-calibration accounts that emphasize experience-based updating rather than smooth linear change at every micro-step of interaction. The analyses involving explanation and context should be interpreted more cautiously. Because explanations were introduced only at the Final moment and the sample size was small, these effects were exploratory. Although the numerical pattern suggests that explanation effectiveness may depend on application context, this result did not reach the conventional significance threshold and requires replication in a larger sample. The study also has several limitations. First, the sample size was small and the study is best understood as a pilot investigation. Second, although EEG was recorded from seven electrodes, the primary confirmatory analysis focused on frontal theta at F3/F4 because this measure was most directly aligned with the theory-driven construct of uncertainty monitoring. Third, questionnaire analyses were limited by the format of the available export and therefore remained descriptive. Future work should use larger samples, report retained trial counts in detail, include broader exploratory analyses across additional frequency bands and electrode groups, and test whether neural and self-report indicators of trust calibration converge.

This pilot NeuroIS study investigated repeated GenAI interaction as a dynamic trust-calibration process and examined frontal theta as a neural correlate of uncertainty monitoring. The results showed a significant reduction in frontal theta from the first AI response to the final decision-relevant moment, providing preliminary evidence that repeated interaction with GenAI may reduce uncertainty monitoring. At the same time, the findings should be interpreted cautiously given the small sample, sparse montage, and exploratory status of the explanation/context analyses. Overall, the study contributes a theory-driven EEG perspective on trust calibration in GenAI interaction and provides a basis for more strongly powered follow-up research.

References

1. Rashid, A.B., Kausik, M.A.K.: AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. Hybrid Advances. 7, 100277 (2024). https://doi.org/10.1016/j.hybadv.2024.100277.
2. Benbasat, I., Wang, W.: Trust In and Adoption of Online Recommendation Agents. JAIS. 6, 72–101 (2005). https://doi.org/10.17705/1jais.00065.
3. Banh, L., Strobel, G.: Generative artificial intelligence. Electron Markets. 33, 63 (2023). https://doi.org/10.1007/s12525-023-00680-1.
4. Krueger, F., Riedl, R., Bartz, J.A., Cook, K.S., Gefen, D., Hancock, P.A., Jarvenpaa, S.L., Krabbendam, L., Lee, M.R., Mayer, R.C., Mislin, A., Müller-Putz, G.R., Simpson, T., Takagishi, H., Van Lange, P.A.M.: A call for transdisciplinary trust research in the artificial intelligence era. Humanit Soc Sci Commun. 12, 1124 (2025). https://doi.org/10.1057/s41599-025-05481-9.
5. Scharowski, N., Perrig, S.A.C., Aeschbach, L.F., Felten, N. von, Opwis, K., Wintersberger, P., Brühlmann, F.: To Trust or Distrust Trust Measures: Validating Questionnaires for Trust in AI, http://arxiv.org/abs/2403.00582, (2025). https://doi.org/10.48550/arXiv.2403.00582.
6. Wang, W., Benbasat, I.: Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents. Journal of Management Information Systems. 33, 744–775 (2016). https://doi.org/10.1080/07421222.2016.1243949.
7. Kalischko, T., Riedl, R.: On the consequences of electronic performance monitoring in organizations: theory and evidence. DTS. 3, 50–79 (2024). https://doi.org/10.1108/DTS-10-2022-0054.
8. Kohn, S.C., De Visser, E.J., Wiese, E., Lee, Y.-C., Shaw, T.H.: Measurement of Trust in Automation: A Narrative Review and Reference Guide. Front. Psychol. 12, 604977 (2021). https://doi.org/10.3389/fpsyg.2021.604977.
9. Li, Y., Wu, B., Huang, Y., Luan, S.: Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. Front. Psychol. 15, 1382693 (2024). https://doi.org/10.3389/fpsyg.2024.1382693.
10. Daronnat, S., Azzopardi, L., Halvey, M., Dubiel, M.: Inferring Trust From Users' Behaviours; Agents' Predictability Positively Affects Trust, Task Performance and Cognitive Load in Human-Agent Real-Time Collaboration. Front. Robot. AI. 8, 642201 (2021). https://doi.org/10.3389/frobt.2021.642201.
11. Alon, L., Levkovich, I.: Trusting the black box: Adapting a multidimensional measure of trust in generative AI. Computers in Human Behavior: Artificial Humans. 8, 100295 (2026). https://doi.org/10.1016/j.chbah.2026.100295.
12. Liao, Q.V., Wortman Vaughan, J.: AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. Harvard Data Science Review. (2024). https://doi.org/10.1162/99608f92.8036d03b.
13. Banh, L., Stangl, F.J., Strobel, G., Riedl, R.: Exploring the NeuroIS Potential for Generative Artificial Intelligence: Findings from a Literature Review.
14. Banh, L., Stangl, F.J., Strobel, G., Riedl, R.: The Role of Generative Artificial Intelligence in the NeuroIS Research Process: Applications and Opportunities.

15. Kalischko, T., Riedl, R.: Physiological Measurement in the Research Field of Electronic Performance Monitoring: Review and a Call for NeuroIS Studies. In: Davis, F.D., Riedl, R., Vom Brocke, J., Léger, P.-M., Randolph, A.B., and Fischer, T. (eds.) Information Systems and Neuroscience. pp. 233–243. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-60073-0_27.
16. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An Integrative Model of Organizational Trust. The Academy of Management Review. 20, 709 (1995). https://doi.org/10.2307/258792.
17. McKnight, D.H., Chervany, N.L.: What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. International Journal of Electronic Commerce. 6, 35–59 (2001). https://doi.org/10.1080/10864415.2001.11044235.
18. Lee, J.D., See, K.A.: Trust in Automation: Designing for Appropriate Reliance. hfes. 46, 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50.30392.
19. Hoff, K.A., Bashir, M.: Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. Hum Factors. 57, 407–434 (2015). https://doi.org/10.1177/0018720814547570.
20. Chiou, E.K., Lee, J.D.: Trusting Automation: Designing for Responsivity and Resilience. Human Factors. 65, 137–165 (2023). https://doi.org/10.1177/00187208211009995.
21. OpenAI: GPT-4 Technical Report. (2023). https://doi.org/10.48550/ARXIV.2303.08774.
22. Liao, Q.V., Vaughan, J.W.: AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. Harvard Data Science Review. (2023). https://doi.org/10.1162/99608f92.8036d03b.
23. Daft, R.L., Lengel, R.H.: Organizational information requirements, media richness and structural design. Management Science. 32, 554–571 (1986). https://doi.org/10.1287/mnsc.32.5.554.
24. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The Role of Trust in Automation Reliance. International Journal of Human-Computer Studies. 58, 697–718 (2003). https://doi.org/10.1016/S1071-5819(03)00038-7.
25. Cavanagh, J.F., Frank, M.J.: Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences. 18, 414–421 (2014). https://doi.org/10.1016/j.tics.2014.04.012.
26. Cavanagh, J.F., Figueroa, C.M., Cohen, M.X., Frank, M.J.: Frontal Theta Reflects Uncertainty and Unexpectedness during Exploration and Exploitation. Cerebral Cortex. 22, 2575–2586 (2012). https://doi.org/10.1093/cercor/bhr332.
27. Cohen, M.X., Donner, T.H.: Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. Journal of Neurophysiology. 110, 2752–2763 (2013). https://doi.org/10.1152/jn.00479.2013.
28. Mas-Herrero, E., Marco-Pallarés, J.: Frontal Theta Oscillatory Activity Is a Common Mechanism for the Computation of Unexpected Outcomes and Learning Rate. Journal of Cognitive Neuroscience. 26, 447–458 (2014). https://doi.org/10.1162/jocn_a_00516.
29. Rescorla, R.A., Wagner, A.R.: A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. Appleton-Century-Crofts, New York (1972).
30. Chen, K.-H., Okerstrom, K.L., Kingyon, J.R., Anderson, S.W., Cavanagh, J.F., Narayanan, N.S.: Startle Habituation and Midfrontal Theta Activity in Parkinson Disease. Journal of Cognitive Neuroscience. 28, 1923–1932 (2016). https://doi.org/10.1162/jocn_a_01012.

31. Muir, B.M.: Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems. Ergonomics. 37, 1905–1922 (1994). https://doi.org/10.1080/00140139408964957.
32. Okamura, K., Yamada, S.: Adaptive Trust Calibration for Human-AI Collaboration. PLOS ONE. 15, e0229132 (2020). https://doi.org/10.1371/journal.pone.0229132.
33. Glikson, E., Woolley, A.W.: Human Trust in Artificial Intelligence: Review of Empirical Research. ANNALS. 14, 627–660 (2020). https://doi.org/10.5465/annals.2018.0057.
34. Burrell, J.: How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society. 3, 2053951715622512 (2016). https://doi.org/10.1177/2053951715622512.
35. Gefen, D., Straub, D.W.: Managing User Trust in B2C e-Services. e-Service Journal. 2, 7–24 (2003). https://doi.org/10.2979/esj.2003.2.2.7.
36. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press (1996).
37. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. Journal of Social Issues. 56, 81–103 (2000). https://doi.org/10.1111/0022-4537.00153.
38. Go, E., Sundar, S.S.: Humanizing Chatbots: The Effects of Visual, Identity and Conversational Cues on Humanness Perceptions. Computers in Human Behavior. 97, 304–316 (2019). https://doi.org/10.1016/j.chb.2019.01.020.
39. Petty, R.E., Cacioppo, J.T.: The Elaboration Likelihood Model of Persuasion. Academic Press (1986). https://doi.org/10.1016/S0065-2601(08)60214-2.
40. Qiu, L., Benbasat, I.: Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. Journal of Management Information Systems. 25, 145–182 (2009). https://doi.org/10.2753/MIS0742-1222250405.
41. Scharowski, N., Perrig, S.A.C., Aeschbach, L.F., Felten, N. von, Opwis, K., Wintersberger, P., Brühlmann, F.: To Trust or Distrust Trust Measures: Validating Questionnaires for Trust in AI. (2025). https://doi.org/10.48550/arXiv.2403.00582.
42. Eisma, J., Rawls, E., Long, S., Mach, R., Lamm, C.: Frontal midline theta differentiates separate cognitive control strategies while still generalizing the need for cognitive control. Sci Rep. 11, 14641 (2021). https://doi.org/10.1038/s41598-021-94162-z.
43. Cavanagh, J.F., Zambrano-Vazquez, L., Allen, J.J.B.: Theta lingua franca: A common mid-frontal substrate for action monitoring processes. Psychophysiology. 49, 220–238 (2012). https://doi.org/10.1111/j.1469-8986.2011.01293.x.
44. Helfrich, R.F., Knight, R.T.: Oscillatory Dynamics of Prefrontal Cognitive Control. Trends in Cognitive Sciences. 20, 916–930 (2016). https://doi.org/10.1016/j.tics.2016.09.007.
45. Cavanagh, J.F., Shackman, A.J.: Frontal Midline Theta Reflects Anxiety and Cognitive Control: Meta-Analytic Evidence. Journal of Physiology-Paris. 109, 3–15 (2015). https://doi.org/10.1016/j.jphysparis.2014.04.003.

Beyond Productivity: Cognitive and Neurophysiological Implications of Generative AI in Knowledge Work

Leonardo Banh[1], Ellen Lachmann[1], and Gero Strobel[1]

[1] University of Duisburg-Essen, Rhine-Ruhr Institute of Information Systems, Essen, Germany
{leonardo.banh,ellen.lachmann,gero.strobel}@uni-due.de

Abstract. The integration of generative AI (GenAI) into knowledge work introduces new forms of collaboration between humans and intelligent systems, raising important questions about how this affects individuals at cognitive and neurophysiological levels. This paper presents a systematic literature review of 39 empirical studies to investigate the psychological and cognitive effects associated with human-GenAI collaboration in knowledge work. Our analysis reveals ten conceptual clusters of individual implications, such as engagement and motivation or self-perception. We further examine how these effects are operationalized and measured, contrasting self-report measures with neurophysiological and behavioral instruments. Overall, the review highlights that current evidence is skewed toward short-term, task-based effects, particularly around cognitive resource dynamics and affective responses, while social, relational, and longitudinal effects remain underexplored. By consolidating the current state of knowledge and mapping how individual effects are operationalized, this review contributes to a structured conceptual foundation for NeuroIS research on human-GenAI collaboration.

Keywords: Generative AI · Knowledge Work · Human-AI Collaboration · NeuroIS · Systematic Literature Review

Introduction

Artificial intelligence (AI) is reshaping the nature of work at a fast pace, moving from a research state and highly specialized fields towards everyday tasks. With the recent popularization of GenAI, AI systems have moved beyond automation into active participation in knowledge work, introducing new forms of collaboration between humans and intelligent systems. By performing tasks that traditionally required human expertise, judgment, or creativity, GenAI is increasingly shaping both knowledge work and creative processes [1]. This development offers opportunities to transform professional as well as personal forms of work [2]. Hence, organizations across various industries explore how the potential of GenAI can be leveraged to enhance knowledge-intensive activities [3, 4]. However, the widespread deployment of GenAI also generates both excitement and concerns regarding possible impact on the workforce [5, 6], particularly regarding potential changes in critical thinking skills and related practices [7, 8].

As GenAI increasingly integrates into knowledge-intensive environments, its influence extends beyond organizational performance to the cognitive and emotional experiences of individual workers [9–11]. Researchers argue that GenAI systems place substantial metacognitive demands on users, requiring them to continuously evaluate and regulate their own cognitive processes [2, 12, 13]. These systems are designed to complement human capabilities, yet they may alter how individuals manage workload, and approach problem-solving [14–17]. For instance, when users delegate analytical or creative tasks to GenAI, they may experience a reduction in mental effort [18, 19]. Moreover, continuous interaction with intelligent systems could affect stress regulation, attention control, and overall mental load [20–22]. This tension between cognitive relief and burden may have measurable effects at the neurophysiological level, reflecting internal cognitive and emotional states [23–26].

However, a coherent understanding of the cognitive and psychological implications associated with GenAI is still nascent [27]. Research and discourse on GenAI engage with its economic potential [5, 28], ethical implications [29–31], impact on education [32, 33] and much more. These perspectives, while valuable, tend to neglect the psychological dimensions of human-GenAI interaction [34]. Yet this human-centered perspective is crucial because the success of GenAI integration depends not only on technical performance but also on how humans perceive and use these systems in their everyday routines [12, 35]. Taking an interdisciplinary view through the NeuroIS perspective that connects insights from neuroscience and information systems allows for a complementary assessment beyond typical IS research to examine cognitive and affective responses during interaction with intelligent systems [36, 37].

Hence, this paper aims to investigate these individual implications from a NeuroIS perspective. By conducting a systematic literature review (SLR), we explore how the integration of GenAI into knowledge work affects individuals at cognitive and neurophysiological levels. By synthesizing existing studies, we examine which individual effects are associated with human-GenAI collaboration and how these effects have been conceptualized and measured. To address this objective, we ask following research question: **Which neurophysiological and cognitive effects are associated with human-GenAI collaboration in knowledge work and how are these measured?**

The remainder of this paper is structured as follows: Section 2 provides the theoretical background on GenAI in knowledge work and the NeuroIS perspective. Section 3 describes the research design. Section 4 presents the results, including the conceptual framework of individual implications and the measurement landscape. Section 5 discusses the findings, and Section 6 concludes with limitations and avenues for future research.

A NeuroIS Lens on GenAI in Knowledge Work

GenAI refers to a class of AI models capable of generating novel content (e.g., text, images, code, audio) by learning the statistical structure and distribution of training data [4, 38]. Unlike narrow, predictive AI systems designed for specific tasks, modern GenAI systems (e.g., OpenAI's ChatGPT or Google's Gemini) enable users to engage in open-ended dialogue, co-generate documents, write code, and synthesize information across various domains [39, 40]. The recent popularization of GenAI tools has created a pivotal moment for knowledge work, especially because of their accessibility.

Knowledge work, broadly defined as work involving the creation, distribution, or application of knowledge [41, 42], is characterized by its cognitive intensity and reliance on human judgment, expertise, and creativity. Introducing GenAI into this domain creates a novel collaborative paradigm in which humans and AI systems increasingly share cognitive labor [1, 43, 44]. This collaboration differs from prior human-tool relationships because GenAI can engage in higher-order cognitive activities instead of automating repetitive tasks, blurring the boundary between human and machine cognition [45–47]. This shift raises critical questions for IS research about what happens to the individual knowledge worker as their cognitive labor is increasingly shared with, delegated to, or mediated by GenAI systems.

The NeuroIS perspective is particularly relevant for studying human-GenAI collaboration because interaction with intelligent systems produces complex cognitive states, such as shifts in attention, mental effort, and affective arousal, that may not be adequately captured through questionnaires alone [48, 49]. Neurophysiological instruments such as electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), eye-tracking, electrodermal activity (EDA) sensors, and electrocardiography (ECG) have been applied in IS research to capture subtle internal states during human-computer interaction [27, 50]. These measures enable researchers to observe cognitive and affective responses in real time, without interrupting the task, and without relying solely on participants' introspective accuracy. Hence, we adopt the NeuroIS perspective both as a theoretical framing, emphasizing cognitive and neurophysiological mechanisms, and as a methodological reference point for evaluating how individual implications of GenAI are measured in the existing literature.

Review Design

This study employs a systematic literature review following established methodological guidelines by Webster & Watson [51] and vom Brocke et al. [52] to investigate the individual implications of using GenAI in knowledge work by focusing on neurophysiological effects. The aim of the review is to identify relevant constructs (e.g., cognitive load, stress, self-efficacy, or overreliance) and analyze how these effects are conceptualized and operationalized in literature through a scoping review [53]. To capture a wide range of literature, our search strategy for peer-reviewed journal articles and conference papers included the academic meta-databases Scopus and Web of Science as well as the domain-specific databases AISeL, ACM Digital Library, and IEEE Xplore. We also expanded our search to all NeuroIS proceedings to integrate further relevant articles. The search was conducted in December 2025 using a search string combining terms related to GenAI (e.g., "generative AI", "large language model", "ChatGPT", "AI assistant"), psychological or cognitive effects (e.g., "cognit*", "load", "stress", "effort"), neurophysiological measurements (e.g., "nervous system", "biometric", "EEG", "heart rate"), and human-interaction contexts (e.g., "collaboration", "interaction").

Initially, 575 publications were identified, which were further reduced to 139 relevant papers based on title/abstract screening and removal of duplicates [54]. During full-text screen, we applied following exclusion criteria: (1) no empirical investigation of individual effects (e.g., conceptual works, no psycho-physiological outcomes); (2) purely technical works without user studies; (3) non-peer-reviewed articles (e.g., comments, opinions); or (4) non-English publications. Inclusion criteria primarily revolved around studies providing evidence on the neurophysiological effects and cognitive mechanisms of GenAI use in work-related contexts, resulting in 32 publications for the final literature sample (see Table 4). Backwards and forwards searches revealed seven additional relevant papers, yielding a total corpus of 39 publications.

These 39 papers were then iteratively analyzed following an open-coding approach to structure the corpus and elicit individual effects mentioned as outcomes of GenAI use and constructs of interest. In line with Webster & Watson [51], higher-order concepts were derived in an axial-coding step that subsumes the initial codes. Moreover, operationalization details were assessed and categorized to provide an overview of how effects were measured. More details regarding the concept matrix and the operationalization can be found in the online appendix (https://bit.ly/4aZSj6A).

Table 4. Literature review process

Database	Initial	Title/Abstract	Full Paper
AIS eLibrary	70	36	12
ACM Digital Library	73	14	5
IEEE Xplore	84	21	4
NeuroIS Proceedings	17	17	5
Scopus	331	55	4
Web of Science	67	9	2
Sum	**575**	**139**	**32**
		Forwards	5
		Backwards	2
		Total	**39**

Results

The iterative open-coding analysis of the relevant papers revealed ten conceptual clusters of individual effects associated with human-GenAI collaboration in knowledge work. Table 5 presents the concept matrix of the reviewed publications, whereas Table 6 provides an overview of all clusters with representative constructs and references.

Table 5. Concept Matrix of Reviewed Publications (*n* = 39)

Study	Attention & Focus	Cognitive Load & Resources	Analytical & Reflective Thinking	Learning & Comprehension	Creativity & Idea Flow	Engagement, Motiva-tion & Agency	Affect & Emotion	Trust, Risk, Privacy & Safety	Self-Perception	Social Embeddedness & Interaction
[55]								✓		✓
[46]	✓							✓		
[56]							✓			✓
[57]	✓									
[58]		✓		✓		✓				
[59]	✓	✓					✓	✓		
[60]	✓	✓					✓			
[61]		✓			✓					
[62]		✓			✓					
[63]		✓		✓			✓			
[64]							✓	✓	✓	✓
[65]		✓								
[66]		✓	✓	✓				✓		
[67]		✓						✓		
[68]		✓	✓							
[69]									✓	
[70]	✓	✓		✓		✓	✓	✓	✓	
[71]						✓		✓		
[72]		✓	✓							
[73]	✓		✓		✓	✓				
[74]			✓							
[75]		✓				✓	✓	✓		✓
[23]		✓				✓			✓	
[76]									✓	
[77]				✓		✓		✓	✓	
[78]							✓			
[24]		✓			✓		✓			

[79]		✓	✓	✓		✓				
[25]		✓					✓			
[80]					✓					
[81]		✓	✓	✓						
[82]		✓	✓	✓						
[83]								✓	✓	
[84]	✓	✓	✓							
[85]		✓								
[86]									✓	
[87]	✓						✓	✓		
[88]						✓				
[89]		✓		✓			✓			
Total	**8**	**22**	**9**	**9**	**5**	**9**	**12**	**12**	**8**	**4**

Individual Implications of GenAI Use in Knowledge Work

The first concept is **Attention & Focus** and encompasses effects that describe how attentive individuals are when working with GenAI to complete a task. In studies, this concept captures constructs such as attention [57, 59], focus [60, 70], or concentration [73, 84]. GenAI tools may enhance focus by absorbing tedious subtasks but may equally also disrupt it by introducing frequent interaction cycles that interrupt sustained cognitive engagement.

Cognitive Load & Resources refers to the mental (and sometimes temporal) effort required to complete a task. Typical constructs investigated include cognitive load [24, 25, 65, 67, 81], workload [59, 60, 86], effort [71, 72], and relief [79, 84]. While GenAI often reduces objective task complexity, the metacognitive demands of prompt engineering and output verification may introduce compensating load.

Analytical & Reflective Thinking captures the depth and quality of reasoning, especially when analyzing data, weighing decisions, or reflecting on AI-generated outputs. Studies investigate critical thinking [66, 81, 84], cognitive analysis [84], cognitive stimulation [73, 74], and self-reflection [68]. A central concern in this cluster is whether continuous use of GenAI for analytical tasks erodes users' own analytical capacity over time [90].

Learning & Comprehension describes how the use of GenAI affects the understanding and learning of content, i.e., how well information is absorbed, processed, and subsequently retained or reproduced. Prior studies examine comprehension [63, 66, 70, 77], learning [79], and information processing [58]. Particularly in educational contexts, the relationship between GenAI assistance and deeper learning outcomes is assessed due to its societal impact [33].

Creativity & Idea Flow summarizes effects describing the influence of GenAI use on idea generation and flexible thinking. For instance, creativity [24, 80], flexibility

[61, 73], or fluency [61, 62] are core constructs of recent works. GenAI can serve as an “idea accelerator” by offering diverse conceptual starting points, yet may also constrain originality by anchoring users to AI-generated frames [3, 14].

Engagement, Motivation & Agency includes effects related to active participation, motivational states, and the sense of control or self-determination when performing tasks in collaboration with GenAI. Constructs include cognitive engagement [58, 70, 77, 84, 88], motivation [75], perceived agency [71], and curiosity [75].

Affect & Emotion subsumes publications investigating emotional reactions when engaging with GenAI, both positive and negative. Here, works investigate outcomes such as satisfaction [70], appreciation and acceptance [78], relief [64, 75], frustration [24, 59, 60], confusion [63], anxiety and restlessness [56, 75], concerns [56], or surprise and astonishment [75] that are caused by human GenAI use and collaboration.

Trust, Risk, Privacy & Safety bundles effects related to trust in GenAI systems and their outputs, as well as perceived risk and security or privacy concerns. The widespread accessibility of GenAI systems such as ChatGPT in private and corporate scenarios poses novel challenges to safe use. Hence, studies examine how GenAI affects trust [67, 71], perceived risk [55], confidentiality concerns [70, 83], or safety needs [64].

Self-Perception encompasses works assessing the effects of GenAI on human abilities and self-image, including self-confidence, self-efficacy, and professional identity. Typical constructs include self-confidence [70, 77, 83], self-esteem [64], self-efficacy [69], or self-actualization [64].

Social Embeddedness & Interaction refers to belonging, relationship experiences, and social interaction in the context of GenAI use. Studies investigate constructs such as belonging [64], attachment [75], social interaction [55], and social level [56].

Table 6. Overview of conceptual clusters, representative constructs, and key references

Conceptual Cluster	Representative Constructs	Key References
Attention & Focus	attention, focus, concentration	[46, 57, 59, 60, 70, 73, 84, 87]
Cognitive Load & Resources	cognitive load, workload, effort, relief	[23–25, 58–63, 65–68, 70, 75, 79, 81–85, 89]
Analytical & Reflective Thinking	critical thinking, cognitive analysis, self-reflection	[66, 68, 72–74, 79, 81, 82, 84]
Learning & Comprehension	comprehension, learning, information processing	[58, 63, 66, 70, 77, 79, 81, 82, 89]
Creativity & Idea Flow	creativity, flexibility, fluency	[24, 61, 62, 73, 80]
Engagement, Motivation & Agency	engagement, motivation, agency, curiosity	[23, 58, 70, 71, 73, 75, 77, 79, 88]
Affect & Emotion	satisfaction, frustration, anxiety, relief, confusion	[24, 56, 59, 60, 63, 64, 70, 75, 78, 80, 87, 89]

Conceptual Cluster	Representative Constructs	Key References
Trust, Risk, Privacy & Safety	trust, perceived risk, confidentiality, safety	[46, 55, 59, 64, 66, 67, 70, 71, 75, 77, 83, 87]
Self-Perception	self-confidence, self-efficacy, self-esteem	[23, 64, 69, 70, 76, 77, 83, 86]
Social Embeddedness & Interaction	belonging, attachment, social interaction	[55, 56, 64, 75]

Operationalization and Measurement

Beyond the analysis of individual implications, our subsequent analysis focused on how these effects are operationalized and measured across the papers. We distinguish three main categories of measurement approaches (see also Table 7).

Self-report measures represent the most prevalent approach in the reviewed corpus, including Likert-scale questionnaires, established psychological scales (e.g., NASA-TLX for workload [91] or the User Engagement Scale (UES) [92]), and open-ended reflective instruments. While ecologically valid and easy to administer, self-reports are subject to biases such as social desirability and retrospective distortion. For instance, a concern in human-GenAI interaction research is that users may be unaware of cognitive changes occurring during task performance, thus making post-hoc reports an imperfect proxy for internal states that call for additional measures at interaction time.

Neurophysiological measures appear in a smaller but growing subset of studies investigating GenAI at the individual level. Instruments include EEG to capture neural oscillations associated with cognitive load and attentional states, fNIRS for prefrontal cortex activation as an index of working memory and executive function, eye-tracking for gaze behavior and pupil dilation as continuous indicators of cognitive effort and focus, electrodermal activity (EDA/GSR) for sympathetic arousal linked to stress and emotional processing, ECG for heart rate variability as a marker of cognitive and emotional regulation, and functional MRI (fMRI) for spatial mapping of neural activation patterns within the brain [27]. *Eye-tracking* is the most frequently applied method in our analyzed set, used to capture fixation duration, pupil dilation, and gaze transitions as indicators of cognitive effort and attentional allocation, for instance to examine how users process GenAI versus human-generated outputs [46] or to detect overreliance on GenAI-generated content in workplace settings [66]. *EEG* constitutes the second most prevalent method, capturing neural activities associated with cognitive load dynamics during AI-assisted writing [24] or cognitive offloading to GenAI in strategic tasks [23]. More recently, *fNIRS* has been introduced to assess prefrontal cortex activation during LLM-assisted concept generation [62] or in combination with EEG to investigate overreliance effects on divergent thinking [84]. Complementary *ANS-based measures* include EDA as a sympathetic arousal indicator [25, 61] and ECG for heart rate variability during GenAI-assisted design ideation [59]. These measures provide continuous access to physiological states during task performance, offering a layer of insight inaccessible to questionnaires alone [36, 37].

Behavioral and performance measures constitute a third category, encompassing task completion time, error rates, output quality ratings, and interaction logs (e.g., number of prompts issued, revision behaviors). These objective indicators complement self-report and neurophysiological data by capturing behavioral patterns of underlying cognitive states. In several studies, behavioral outcomes (e.g., code correctness, text quality scores) are used as proxies for cognitive performance while subjective measures capture perceived experience [72, 74, 78].

Table 7. Distribution of measurement approaches in the reviewed corpus

Measurement Approach	Studies	Example Instruments
Self-report	20 (59%)	NASA-TLX, UES, questionnaires, interviews
Neurophysiological	16 (33%)	EEG, fNIRS, eye-tracking, EDA, ECG, fMRI
Behavioral/performance	3 (8%)	Task completion time, error rates, output quality

Discussion and Conclusion

This SLR synthesizes 39 empirical studies on the individual implications of human-GenAI collaboration in knowledge work and reveals a field that is rapidly expanding yet conceptually and methodologically scattered. Across the corpus, we synthesized ten conceptual clusters that capture a broad spectrum of cognitive, affective, and social constructs. However, research attention is mainly focused on *Cognitive Load & Resources* and *Affect & Emotion*. In contrast, *Social Embeddedness & Interaction*, *Trust/Risk/Privacy & Safety*, and *Self-Perception* remain underrepresented within the human-GenAI collaboration context.

Our analysis of operationalizations highlights that there is a reliance on self-report measures in current studies, calling for a broader methodological approach to assessing human-GenAI collaboration in favor of NeuroIS and mixed-method studies [24, 46, 66]. While NeuroIS methods such as EEG and eye-tracking were already employed to investigate GenAI implications from a neurophysiological perspective, other methods including fMRI or hormone-based biomarkers such as cortisol and oxytocin were absent from the reviewed corpus, despite their established value in broader NeuroIS research [27, 36]. Moreover, current works are mostly short-term and task-based as they assess immediate post-task effects in tightly controlled settings (e.g., writing, coding, or information-search tasks). This research design highlights performance-contingent outcomes but leaves cumulative or developmental consequences on the human individual unexplored [11]. Whether repeated GenAI delegation fosters cognitive overreliance, decreases critical thinking, or reshapes professional identity remains questions worthwhile pursuing, especially with (generative) AI evolving quickly [8, 23, 82, 90, 93]. Longitudinal and field-based designs are therefore essential to move beyond momentary snapshots toward processual understanding [75]. Finally, conceptual clarity remains limited. Constructs such as engagement, attention, cognitive load, mental effort, and workload are sometimes used interchangeably or defined inconsistently [23, 87]. This fragmentation makes cumulative knowledge building more difficult. Future research would benefit from clearer construct definitions, validated measurement instruments tailored to human-GenAI contexts, and stronger grounding in established cognitive, motivational, and social theories.

Yet this study is not without limitations. Despite a comprehensive search strategy, relevant studies may have been missed due to the selection criteria and individual screening process, particularly those works examining psychological outcomes without explicit neurophysiological terminology or older recent works focusing on agentic AI systems [94] without referring to GenAI. The clustering and coding process involves interpretive judgment and theoretical positioning, despite iterations within the author team. Taken together, this review contributes to NeuroIS and the broader IS literature by offering (1) a structured conceptual map of individual effects of human-GenAI collaboration and (2) a systematic overview of how these effects are operationalized.

As GenAI systems increasingly diffuse into professional and private knowledge work, understanding their individual implications remains important both on a scholarly and practical point of view. Organizations require evidence-based guidance to deploy these technologies responsibly, while researchers must develop robust frameworks to assess how GenAI shapes cognition, affect, identity, and social interaction over time. By consolidating the current state of knowledge and articulating a forward-looking agenda, our review aims to provide a foundation for cumulative, theory-driven, and methodologically sophisticated research on human-GenAI collaboration.

Acknowledgements. This work was supported by the Stiftung Innovation in der Hochschullehre (Project No. 3004-0453).

References

1. Benbya, H., Strich, F., Tamm, T.: Navigating Generative Artificial Intelligence Promises and Perils for Knowledge and Creative Work. JAIS, vol. 25, 23–36 (2024). doi: 10.17705/1jais.00861
2. Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A.E., Sarkar, A., Sellen, A., Rintel, S.: The Metacognitive Demands and Opportunities of Generative AI. In: Proceedings of the CHI '24: CHI Conference on Human Factors in Computing Systems, pp. 1–24. ACM (2024). doi: 10.1145/3613904.3642902
3. Grange, C., Demazure, T., Ringeval, M., Bourdeau, S., Martineau, C.: The Human-GenAI Value Loop in Human-Centered Innovation: Beyond the Magical Narrative. Information Systems Journal, vol. 36, 29-51 (2025). doi: 10.1111/isj.12602
4. Banh, L., Strobel, G.: Generative artificial intelligence. Electron Markets, vol. 33 (2023). doi: 10.1007/s12525-023-00680-1
5. Brynjolfsson, E., Li, D., Raymond, L.: Generative AI at Work. The Quarterly Journal of Economics, vol. 140, 889–942 (2025). doi: 10.1093/qje/qjae044
6. Noy, S., Zhang, W.: Experimental evidence on the productivity effects of generative artificial intelligence. Science, vol. 381, 187–192 (2023). doi: 10.1126/science.adh2586

7. Korosec-Serfaty, M., Léger, P.-M., Parent-Rocheleau, X., Sénécal, S.: Critical Thinking in AI-Assisted Deontologically-Governed Professional Decision-Making: When and How Explainability, Reliability, and Transparency Matter. Information Systems Frontiers, vol. (2026). doi: 10.1007/s10796-025-10691-2
8. Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., Wilson, N.: The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. In: Proceedings of the CHI 2025: CHI Conference on Human Factors in Computing Systems, pp. 1–22. ACM (2025). doi: 10.1145/3706598.3713778
9. Jia, M., Zhao, Y.C., Zhang, X.: Will potential threats persist? Attitudes and behavioral responses of creative professionals toward GenAI. Information Research an international electronic journal, vol. 30, 1073–1083 (2025). doi: 10.47989/ir30iConf47098
10. Law, M., Varanasi, R.A., Siau, K.L., Nah, F.F.-H.: Generative AI and Changing Work: Systematic Review of Practitioner-Led Work Transformations Through the Lens of Job Crafting, vol. 15804, 131–152 (2025). doi: 10.1007/978-3-031-92823-9_10
11. Wu, S., Liu, Y., Ruan, M., Chen, S., Xie, X.-Y.: Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Sci Rep, vol. 15, 15105 (2025). doi: 10.1038/s41598-025-98385-2
12. Lee, Y., Ahn, H., Jo, H.: Exploring the Drivers and Outcomes of Generative AI Usage: The Role of Metacognition. International Journal of Human–Computer Interaction, vol. 42, 2098–2120 (2026). doi: 10.1080/10447318.2025.2526592
13. Zahn, M. von, Liebich, L., Jussupow, E., Hinz, O., Bauer, K.: Knowing (Not) to Know: Explainable Artificial Intelligence and Human Metacognition. Information Systems Research, vol. 0 (2025). doi: 10.1287/isre.2024.1431
14. Boussioux, L., Lane, J.N., Zhang, M., Jacimovic, V., Lakhani, K.R.: The Crowdless Future? Generative AI and Creative Problem-Solving. Org. Sci., vol. 35, 1589–1607 (2024). doi: 10.1287/orsc.2023.18430
15. Chen, Z., Chan, J.: Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise. Mang. Sci., vol. 70, 9101–9117 (2024). doi: 10.1287/mnsc.2023.03014
16. Johnson, J.G., Peralta, M., Kaur, M., Huang, R.S., Zhao, S., Guan, R., Rajaram, S., Nebeling, M.: Exploring Collaborative GenAI Agents in Synchronous Group Settings: Eliciting Team Perceptions and Design Considerations for the Future of Work. Proc. ACM Hum.-Comput. Interact., vol. 9, 1–33 (2025). doi: 10.1145/3757595
17. Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X.-H., Beresnitzky, A.V., Braunstein, I. and Maes, P.: Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv, vol. (2025)
18. Shukla, P., Bui, P., Levy, S.S., Kowalski, M., Baigelenov, A., Parsons, P.: De-skilling, Cognitive Offloading, and Misplaced Responsibilities: Potential Ironies of AI-Assisted Design. In: Yamashita, N., Evers, V., Yatani, K., Ding, X. (eds.)

Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–7. ACM, New York, NY, USA (2025). doi: 10.1145/3706599.3719931
19. Shum, S.B.: Generative AI for Critical Analysis: Practical Tools, Cognitive Offloading and Human Agency, vol. (2024)
20. Hessari, H., Bai, A., Daneshmandi, F.: Generative AI: Boosting Adaptability and Reducing Workplace Overload. Journal of Computer Information Systems, vol. , 1–14 (2024). doi: 10.1080/08874417.2024.2417672
21. Hinss, M.F., Brock, A.M., Roy, R.N.: Cognitive effects of prolonged continuous human-machine interaction: The case for mental state-based adaptive interfaces. Front. Neuroergonomics, vol. 3 (2022). doi: 10.3389/fnrgo.2022.935092
22. Yu, P., Zhang, X., Ma, L., Liang, Y.: How Challenge Stressors Affect GenAI-Enabled Employee Innovation? Based on Transactional Theory of Stress and Coping Model. IEEE Transactions on Engineering Management, vol. 72, 2600–2610 (2025). doi: 10.1109/TEM.2025.3578623
23. Ritz, E., Freise, L.R., Li, M.M.: Offloading to Digital Minds: How Generative AI Can Help to Craft Jobs. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G.R. (eds.) Information Systems and Neuroscience, vol. 66, pp. 11–20. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-031-71385-9_2
24. Schulz, T., Knierim, M.: Cognitive Load Dynamics in Generative AI-Assistance: A NeuroIS Study. Proceedings of the International Conference on Information Systems 2024, vol. (2024)
25. Song, X., Song, H., Gao, B., Gao, Q., Li, W., Xu, L., Lu, Z.: “Do You Need the Sage's Tea or the Friend's Cola” Exploring the Differential Healing Effects of Generative AI Conversational Styles. In: Yamashita, N., Evers, V., Yatani, K., Ding, X. (eds.) Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–8. ACM, New York, NY, USA (2025). doi: 10.1145/3706599.3720026
26. Yuan, Q., Yu, M., Chen, D., Cao, Y., Wen, Q., Shen, M., Lin, H.: The Effect of Static and Dynamic AI-Generated Visual Content on Emotional Recall in the Elderly, vol. 15821, pp. 89–106 (2025). doi: 10.1007/978-3-031-93418-6_7
27. Banh, L., Stangl, F.J., Strobel, G., Riedl, R.: Exploring the NeuroIS Potential for Generative Artificial Intelligence: Findings from a Literature Review. Proceedings of the 17th NeuroIS Retreat, vol. (2025)
28. Orchard, T., Tasiemski, L.: The rise of generative AI and possible effects on the economy. Economics and Business Review, vol. 9, 9 (2023). doi: 10.18559/ebr.2023.2.732
29. Georgiou, G.P.: Mapping The Ethical Discourse in Generative Artificial Intelligence: A Topic Modeling Analysis of Scholarly Communication. Language, Technology, and Social Media, vol. 3, 250–265 (2025). doi: 10.70211/ltsm.v3i2.230
30. Kirova, V., Ku, C., Laracy, J., Marlowe, T.: The Ethics of Artificial Intelligence in the Era of Generative AI. Journal of Systemics, Cybernetics and Informatics, vol. 21, 42–50 (2023). doi: 10.54808/JSCI.21.04.42

31. Tortora, L.: Beyond Discrimination: Generative AI Applications and Ethical Challenges in Forensic Psychiatry. Front. Psychiatry, vol. 15 (2024). doi: 10.3389/fpsyt.2024.1346059
32. Mashayekhy, M., Nosrati, F., Ghasemaghaei, M.: Unlocking the Future of Education: Empirical Insights into the Adoption of Generative AI in Higher Education. CAIS, vol. 57, 1270–1295 (2025). doi: 10.17705/1CAIS.05755
33. van Slyke, C., Johnson, R., Sarabadani, J.: Generative Artificial Intelligence in Information Systems Education: Challenges, Consequences, and Responses. CAIS, vol. 53, 1–21 (2023). doi: 10.17705/1CAIS.05301
34. Mazarakis, A., Bernhard-Skala, C., Braun, M., Peters, I.: What is critical for human-centered AI at work? – Toward an interdisciplinary theory. Front. Artif. Intell., vol. 6 (2023). doi: 10.3389/frai.2023.1257057
35. Böhme, C., Graf-Pfohl, C. (eds.): Enhancing Technology Acceptance in Socio-Technical Systems: A Human-Centered Approach to AI Implementation, vol. (2025). doi: 10.54941/ahfe1005842
36. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS. Information Systems and the Brain. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
37. Dimoka, A., Davis, F.D., Gupta, Pavlou, Banker, R.D., Dennis, A.R., Ischebeck, Müller-Putz, G.R., Benbasat, I., Gefen, et al.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly, vol. 36, 679 (2012). doi: 10.2307/41703475
38. Hinds, P., Krogh, G. von: Generative AI, Emerging Technology, and Organizing: Towards a theory of progressive encapsulation. Organization Theory, vol. 5 (2024). doi: 10.1177/26317877241293478
39. Alavi, M., Leidner, D., Mousavi, R.: Knowledge Management Perspective of Generative Artificial Intelligence (GenAI). SSRN Journal, vol. (2024). doi: 10.2139/ssrn.4782875
40. Banh, L., Holldack, F., Strobel, G.: Copiloting the future: How generative AI transforms Software Engineering. Information and Software Technology, vol. 183, 107751 (2025). doi: 10.1016/j.infsof.2025.107751
41. Blackler, F.: Knowledge, Knowledge Work and Organizations: An Overview and Interpretation. Organization Studies, vol. 16, 1021–1046 (1995). doi: 10.1177/017084069501600605
42. Iivari, J., Linger, H.: Characterizing Knowledge Work: A Theoretical Perspective. AMCIS 2000 Proceedings., vol. (2000)
43. Dell'Acqua, F., McFowland, E., Mollick, E.R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R.: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School Working Paper, vol. (2023). doi: 10.2139/ssrn.4573321
44. Hernández-Ramírez, R., Ferreira, J.B.: The Future End of Design Work: A Critical Overview of Managerialism, Generative AI, and the Nature of Knowledge Work, and Why Craft Remains Relevant. She Ji: The Journal of Design, Economics, and Innovation, vol. 10, 414–440 (2024). doi: 10.1016/j.sheji.2024.11.002

45. Söllner, M., Arnold, T., Benlian, A., Bretschneider, U., Knight, C., Ohly, S., Rudkowski, L., Schreiber, G., Wendt, D.: ChatGPT and Beyond: Exploring the Responsible Use of Generative AI in the Workplace. Bus Inf Syst Eng, vol. (2025). doi: 10.1007/s12599-025-00932-8
46. Bera, P., Palankar, N.: How Do We Process GenAI Outputs? Evidence From an Eye-Tracking Study. SIGHCI 2024 Proceedings, vol. (2025)
47. Grinschgl, S., Neubauer, A.C.: Supporting Cognition With Modern Technology: Distributed Cognition Today and in an AI-Enhanced Future. Front. Artif. Intell., vol. 5, 908261 (2022). doi: 10.3389/frai.2022.908261
48. Riedl, R., Stangl, F.J.: NeuroIS. In: Galliers, R.D., Stein, M.-K., Baiyere, A. (eds.) The Routledge Companion to Management Information Systems. Routledge, London, UK (2025)
49. Dimoka, A., Pavlou, P.A., Davis, F.D.: Research Commentary —NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. Information Systems Research, vol. 22, 687–702 (2011). doi: 10.1287/isre.1100.0284
50. Riedl, R., Fischer, T., Léger, P.-M., Davis, F.D.: A Decade of NeuroIS Research. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, vol. 51, 13–54 (2020). doi: 10.1145/3410977.3410980
51. Webster, J., Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, vol. 26, xiii–xxiii (2002)
52. vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. Proceedings of the European Conference on Information Systems 2009, vol. (2009)
53. Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S.: Synthesizing information systems knowledge: A typology of literature reviews. Inf. Man., vol. 52, 183–199 (2015). doi: 10.1016/j.im.2014.08.008
54. Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., Beekhuyzen, J.: Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support. CAIS, vol. 37 (2015). doi: 10.17705/1CAIS.03708
55. Albahhar, A., Obasi, C., Jaiteh, E., Javed, I., Mualana, A., Abbasi, G.: The Impact of Enablers and Disablers on Students' Switching Intention from Generative Artificial Intelligent Chatbots. SaudiCIS 2024 Proceedings, vol. (2024)
56. Bruhin, O.: Beyond Code: The Impact of Generative AI on Work Systems in Software Engineering. ICIS 2024 Proceedings, vol. (2024)
57. Coban, A., Dzsotjan, D., Küchemann, S., Durst, J., Kuhn, J., Hoyer, C.: AI support meets AR visualization for Alice and Bob: personalized learning based on individual ChatGPT feedback in an AR quantum cryptography experiment for physics lab courses. EPJ QUANTUM TECHNOLOGY, vol. 12 (2025). doi: 10.1140/epjqt/s40507-025-00310-z
58. Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., Abrahamson, D.: Generative AI and multimodal data for educational feedback: Insights from embodied math learning. British Journal of Educational Technology, vol. 56, 1686–1709 (2025). doi: 10.1111/bjet.13587

59. Fu, J., Zhong, W., Shen, M., Chen, D.: Evaluating the impact of AIGC-Supported design ideation on Designers' cognitive load and creativity. Displays, vol. 91 (2026). doi: 10.1016/j.displa.2025.103275
60. Gadhvi, R., Petkar, S., Desai, P.P., Ramachandran, S., Siddharth, S.: AdaptAI: A Personalized Solution to Sense Your Stress, Fix Your Mess, and Boost Productivity. In: Yamashita, N., Evers, V., Yatani, K., Ding, X. (eds.) Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–12. ACM, New York, NY, USA (2025). doi: 10.1145/3706599.3720284
61. Ge, S., Sun, Y., Cui, Y., Wei, D.: An Innovative Solution to Design Problems: Applying the Chain-of-Thought Technique to Integrate LLM-Based Agents With Concept Generation Methods. IEEE Access, vol. 13, 10499–10512 (2025). doi: 10.1109/ACCESS.2024.3494054
62. Ge, S., Huang, S., Song, X., Sun, Y., Wang, N., Qi, K., Zhao, X., Li, J., Cui, Y.: Comparing TRIZ and brainstorming in human-agent design collaboration: effects on cognitive processes and performance. Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM, vol. 39 (2025). doi: 10.1017/S0890060425100097
63. Higasa, T., Tanaka, K., Feng, Q., Morishima, S.: Keep Eyes on the Sentence: An Interactive Sentence Simplification System for English Learners Based on Eye Tracking and Large Language Models Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1–7. Association for Computing Machinery, Honolulu, HI, USA (2024). doi: 10.1145/3613905.3650792
64. Hsu, H.-P.: An autoethnographic study of ESL academic writing with ChatGPT: from psychological insights to the SUPER framework. Cogent Education, vol. 12 (2025). doi: 10.1080/2331186X.2025.2543113
65. Knierim, M.T., Del Puppo, L.: Do Generative AI Tools Foster Positive Experiences in Knowledge Work? A NeuroIS Research Proposal. 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 2023, Vienna, Austria. Lecture Notes in Information Systems and Organization, vol. 68, pp. 57–65 (2023). doi: 10.1007/978-3-031-58396-4_6
66. Langner, M., Toreini, P., Maedche, A.: Eye-Tracking Technology for Cognition-Adaptive LLM-Based Assistants at the Workplace. In: Coursaris, C.K., Beringer, J., Léger, P.-M., Öz, B. (eds.) The Design of Human-Centered Artificial Intelligence for the Workplace. Progress in IS, pp. 293–303. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-031-83512-4_17
67. Leheta, M., Ruiz-Segura, A., Karran, A.J., Coursaris, C., Charland, P., Sénécal, S., Léger, P.-M.: Evaluating Voice-Based AI for Grammar Learning: Does Richer Media Improve Learning? Lecture Notes in Information Systems and Organization, vol. , 317–326 (2025)
68. Liu, S., Mädche, A., Feick, M.: GazeClass: Towards Gaze-Adaptive Cross-Device Learning Support for Virtual Classrooms. In: Yamashita, N., Evers, V., Yatani, K., Ding, X. (eds.) Proceedings of the Extended Abstracts of the CHI

Conference on Human Factors in Computing Systems, pp. 1–12. ACM, New York, NY, USA (2025). doi: 10.1145/3706599.3720232
69. Ma, Y., Wang, Z., Pang, H.: The more the merrier? Examining the effects of a conversational agent on EFL learners' speaking in three conditions. Computers & Education, vol. 239 (2025). doi: 10.1016/j.compedu.2025.105442
70. Mahmoudi, M., Taghiyareh, F., Hessami, F.: Enhancing Learning Performance through LLM-Driven Adaptive Contents Based on Facial Engagement Detection. In: 2025 29th International Computer Conference, Computer Society of Iran (CSICC), pp. 1–7. IEEE (2025). doi: 10.1109/CSICC65765.2025.10967446
71. Malaguti, P., Karran, A., Di Le, Mortin, H., Coursaris, C., Sénécal, S., Léger, P.-M.: Investigating Interaction Friction in Generative AI: Improving User Experience and Decision-Making. SIGHCI 2024 Proceedings, vol. (2025)
72. Mehler, M., Ellenrieder, S., Buxmann, P.: The Influence of Effort on the Perceived Value of Generative AI: A Study of the IKEA Effect. ECIS 2024 Proceedings, vol. (2024)
73. Memmert, L., Bittner, E.: Human-AI Collaboration for Brainstorming: Effect of the Presence of AI Ideas on Breadth of Exploration. Hawaii International Conference on System Sciences 2024 (HICSS-57), vol. (2024)
74. Memmert, L.: Brainstorming with a Generative Language Model: Understanding Performance Through Brainstorming Group Effects. ECIS 2024 Proceedings, vol. (2024)
75. Retkowsky, J., Hafermalz, E., Huysman, M.: From playmate to assistant; User experiences of integrating ChatGPT into knowledge work. ICIS 2023 Proceedings, vol. (2023)
76. Sammon, D., McCarthy, S., Thummadi, B.V., Wibisono, A., Fitzgerald, B.: Exploring the Potential of Large Language Models (LLMs) for Grounded Theorizing: A Human-in-the-Loop Configuration. ICIS 2024 Proceedings, vol. (2024)
77. Santhosh, J., Dengel, A., Ishimaru, S.: Gaze-Driven Adaptive Learning System With ChatGPT-Generated Summaries. IEEE Access, vol. 12, 173714–173733 (2024). doi: 10.1109/ACCESS.2024.3503059
78. Schmidt, C., Guffler, M., Kindermann, B., Flatten, T.: Collaborating with Generative AI: Exploring Algorithm Appreciation in Creative Writing. ICIS 2023 Proceedings, vol. (2023)
79. Schulz, T., Krisam, C., Seitz, J.: EyeGPT: A Cognitive Load-Adaptive GenAI Assistant with Eye Tracking for Programming Education. 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 2025, Vienna, Austria. Lecture Notes in Information Systems and Organisation, vol. 9, pp. 45–55. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-032-00815-2_4
80. Song, X., Xu, L., Peng, C., Pan, S., Adibi, A., Wang, X., Lu, Z.: Enhanced creativity at the cost of increased stress? The impact of generative AI on serious games for creativity stimulation. Behaviour and Information Technology, vol. (2025). doi: 10.1080/0144929X.2025.2459267
81. Sqalli, M.T.: Eyes on the Code: Mapping Critical Thinking Through Eye-Tracking for Student-LLM Coding Interactions. In: Proceedings of the 16th Biannual

Conference of the Italian SIGCHI Chapter, pp. 1–13. ACM, New York, NY, USA (2025). doi: 10.1145/3750069.3750397
82. Stadler, M., Bannert, M., Sailer, M.: Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. Computers in Human Behavior, vol. 160, 108386 (2024). doi: 10.1016/j.chb.2024.108386
83. Svensson, B., Keller, C.: Agentic Relationship Dynamics in Human-AI Collaboration: A study of interactions with GPT-based agentic IS artifacts. Hawaii International Conference on System Sciences 2024 (HICSS-57), vol. (2024)
84. Tadson, B., Krisam, C., Mädche, A.: Understanding Overreliance in Human-GenAI Interactions and Its Impact on Divergent Thinking: A NeuroIS Study. 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 2025, Vienna, Austria. Lecture Notes in Information Systems and Organisation, pp. 143–157. Springer Nature Switzerland, Cham (2025). doi: 10.1007/978-3-032-00815-2_14
85. Tang, N., Chen, M., Ning, Z., Bansal, A., Huang, Y., McMillan, C., Li, T.J.-J.: Developer Behaviors in Validating and Repairing LLM-Generated Code Using IDE and Eye Tracking. In: 2024 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 40–46. IEEE (2024). doi: 10.1109/VL/HCC60511.2024.00015
86. Taudien, A., Fügener, A., Gupta, A., Ketter, W.: The Effect of AI Advice on Human Confidence in Decision-Making. Hawaii International Conference on System Sciences 2022 (HICSS-55), vol. (2022)
87. Todd, T.N., Lakhiwal, A., Bala, H., Léger, P.-M., Kamouni, I.E., Boasen, J.: Generative AI Review Summaries and Their Impact on Ambivalence and User Behavior: An Eye-Tracking Study. 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 2024, Vienna, Austria. Lecture Notes in Information Systems and Organization (2024)
88. Yatani, K., Sramek, Z. and Yang, C.-L.: AI as Extraherics: Fostering Higher-order Thinking Skills in Human-AI Interaction. arXiv, vol. (2024)
89. Zou, B., Wang, C., He, H., Li, C., Purwanto, E., Wang, P.: Enhancing EFL writing with visualised GenAI feedback: A cognitive affective theory of learning perspective on revision quality, emotional response, and human-computer interaction. Learning and Motivation, vol. 91, 102158 (2025). doi: 10.1016/j.lmot.2025.102158
90. Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W., Ruissalo, J.: The Vicious Circles of Skill Erosion: A Case Study of Cognitive Automation. JAIS, vol. 24, 1378–1412 (2023). doi: 10.17705/1jais.00829
91. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload. Advances in Psychology, vol. 52, pp. 139–183. Elsevier (1988). doi: 10.1016/S0166-4115(08)62386-9
92. O'Brien, H.L., Cairns, P., Hall, M.: A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short

form. International Journal of Human-Computer Studies, vol. 112, 28–39 (2018). doi: 10.1016/j.ijhcs.2018.01.004
93. Chalmers, D., Hunt, R.‘., Pachidi, S., Potočnik, K., Townsend, D.: The Acceleration of Artificial Intelligence: Rethinking Organization and Work in an Era of Rapid Technological Change. J Management Studies, vol. 63, 285–314 (2026). doi: 10.1111/joms.70063
94. Holldack, F., Banh, L., Strobel, G.: Agentic information systems. Electron Markets, vol. 36 (2026). doi: 10.1007/s12525-025-00861-0

Exploring the Impact of the Gestalt Laws of Proximity and Similarity: A NeuroIS Study in the E-Commerce Context

Anita Vogl[1], Laura Königseder[1] and René Riedl[1,2]

[1] Digital Business Institute, School of Business and Management,
University of Applied Sciences Upper Austria, Steyr, Austria
{anita.vogl,rene.riedl}@fh-steyr.at
laura-koenigseder@gmx.at
[2] Institute of Business Informatics – Information Engineering, Johannes Kepler University
Linz, Linz, Austria
rene.riedl@jku.at

Abstract. Gestalt laws such as proximity and similarity shape visual perception and are widely applied in interface design; however, their effects on neurophysiological and emotional processes during human-computer interaction remain insufficiently understood. Drawing on Norman's Emotional Design Framework, this study examines how proximity and similarity influence users' visual processing efficiency (eye-tracking), emotional responses (facial expression analysis and galvanic skin response), and user experience (self-report, UEQ-S) in an e-commerce context. We will conduct a 2 × 2 within-subject e-commerce experiment manipulating proximity and similarity. The findings contribute to a better understanding of how fundamental principles of perceptual organization influence autonomic emotional processes and subjective feelings. These insights are expected to improve user interface design.

Keywords: User experience (UX) design · Gestalt law · proximity · similarity · neuroscientific methods · eye-tracking study · galvanic skin response · facial recognition · brain · autonomic nervous system · interface design.

Introduction

Gestalt laws are commonly used in user experience design, either for visual, auditive, or haptic input and output. The concept of Gestalt psychology and the related theory was introduced already in 1890 [8] and describes how humans naturally perceive and organize visual elements into unified wholes based on patterns, proximity, similarity, continuity, closure, connectedness, and some other factors. With the continuous technological advancement and the growing number of digital systems, users increasingly interact with a wide variety of user interfaces in various contexts. This underscores the growing significance of intuitive interfaces. Gestalt laws play a fundamental role in shaping how users organize, interpret, and understand information. Based on this, design patterns have been developed and applied to user interfaces [12]. These patterns aim to enhance clarity, usability, and visual understanding by aligning interface structures with human perceptual processes.

Moreover, Norman presented the concept of emotional design as early as 2004 [16], arguing that designers need to consider the users' reflective, behavioral, and visceral responses to design. Visceral responses are grounded in unconscious autonomic processes. This perspective aligns with the goal of enhanced user experience [10]. As Gestalt laws are based on the holistic and configural properties, defined by the interrelations of components, their application has an influence on the perception of user interfaces and their usability. However, it remains unclear whether these factors also affect the visceral response and how these responses interact with behavioral and reflective responses.

To better understand Gestalt laws, previous experiments have focused on user interface or website design such as enhancing the current understanding of the perceptual features that favor users' interactions with websites [5], studying the influence of Gestalt laws and visualization in website design on the degree of acceptance and recommendation [14], and studying the role of perceived user interface quality [7]. In the E-Commerce context, research examined the attention to distracting products unrelated to the shopping goal [11] and investigated the effects of color and shapes on e-commerce websites [15]. Moreover, data and information visualization experiments focused on studying the effects of Gestalt laws on viewing behavior and user performance [3] and how proximity of data elements in social data charts influences viewers' perceptions of unity among the data points [18]. Recent studies also discussed how colors can be used to counteract the decline in visual perception [13] and examined mouse movements as interaction technique to solve tasks in simple visual interface settings [6].

However, little is known about how the perceptual and aesthetic features of user interfaces based on the application of Gestalt laws influence the emotions related to cognitive effort, perceived usability, and perceived user experience. Generally, previous research on Gestalt laws has focused on mainstream self-report usability metrics like learnability, ease of use or usefulness [7, 19], eye-tracking measures like fixation duration or fixation count as well as heatmaps for visual confirmation of results [2, 3]. Also, research examined attention and social presence [11, 14], as well as user ratings and acceptance evaluations [14, 18]. Importantly, a few studies using EEG [13], fMRI [17], fNIRS [15], existing research has hardly used neurophysiological measures to better understand the user experience and the users' autonomic reactions to user interfaces. Despite their widespread use in user interface design, the mechanisms through which Gestalt laws influence user experience remain insufficiently understood.

In this paper, we present an experiment investigating the influence of the Gestalt laws proximity and similarity in two-dimensional visual user interfaces on users' emotions, visual processing behavior and cognitive effort as well as on perceived usability and user experience. By integrating physiological, behavioral, and subjective measures, this study provides a more comprehensive understanding of how fundamental perceptual design principles influence both unconscious neurophysiological processes and conscious evaluations, thereby contributing to NeuroIS research and informing the design of more efficient and emotionally engaging user interfaces.

Our work is focused on the following research questions – based on the following logic of the causal chain: Gestalt laws → visual processing (eye behavior) → emotional response (visceral) → user experience (reflective based on self-report):

RQ1: How does the application of the Gestalt laws of proximity and similarity affect users' visual processing behavior and cognitive effort during interface interaction?

RQ2: How does the application of the Gestalt laws of proximity and similarity influence users' emotions as reflected in arousal and valence

RQ3: How does the application of the Gestalt laws of proximity and similarity influence users' perceived usability and user experience (measured by pragmatic and hedonic quality)?

This paper is structured as follows: In Section 0, we outline the proposed research model. In Section 0, we describe the methodology, including the experimental design, participants, procedure and tasks, and the planned data collection. In Section 0, we discuss the current limitations and in Section 0, we summarize the proposed experiment.

Theoretical Foundation and Research Model

The design of user interfaces fundamentally shapes how users perceive, process, and emotionally respond to user interfaces. According to Norman's theory of emotional design, human interaction with products operates across three interconnected levels of processing: the visceral, behavioral, and reflective levels [16]. These levels represent distinct but interrelated stages through which users automatically perceive visual stimuli, interact with systems, and form conscious evaluations of their experience. Gestalt laws such as proximity and similarity directly influence early perceptual organization and are therefore particularly relevant for understanding how interface design affects visual processing, emotional responses, and overall user experience.

Visual processing behavior and cognitive effort: Norman [16] emphasizes that on the behavioral level humans act based on their well-learned and routine operations. The human perceptual system is highly sensitive to structural organization, and users rely on visual cues to interpret relationships between interface elements. Gestalt laws such as proximity and similarity facilitate perceptual grouping by enabling users to quickly identify related elements and distinguish between functional units. This perceptual organization reduces ambiguity and supports efficient interpretation of visual information and subsequently better understandability and usability.

When interface elements are organized according to Gestalt laws, users can more easily form accurate mental representations of the interface structure. Norman [16] argues that well-designed products focus on understanding and satisfying needs. Therefore, good behavioral design reduces cognitive effort required to understand how information is organized and how actions should be performed. Conversely, poorly organized interfaces increase cognitive effort because users must expend additional mental resources to interpret visual relationships and locate relevant information. Efficient perceptual organization therefore reduces the need for extensive visual search and cognitive processing, resulting in more efficient visual processing behavior. In contrast, interfaces that violate Gestalt laws may increase cognitive effort and lead to less efficient visual exploration.

Emotional responses and autonomic reactions: In addition to influencing perceptual organization, visual interface design also affects users' emotional responses. Norman [16] describes that visceral design is all about immediate emotional impact, where sensory input automatically triggers affective reactions before conscious interpretation occurs. These visceral responses are rapid, automatic, and closely linked to the autonomic nervous system. Visual properties such as clarity, coherence, and organization, or just the shape or form, can evoke immediate positive or negative affective responses.

Interfaces that are perceptually well-organized according to Gestalt laws may be processed more fluently and with less effort, which can lead to more positive emotional responses. Norman [16] emphasizes that users are highly sensitive to visual coherence and that perceptually pleasing and well-organized designs tend to evoke more positive affective reactions. In contrast, disorganized or visually confusing interfaces may induce negative affective responses due to increased perceptual and cognitive demands. These emotional reactions can manifest not only in subjective evaluations, but also in measurable physiological responses, reflecting the close connection between perceptual processing and emotional systems. This is important in product or user experience design as we strive for positive emotions [10]. Therefore, we argue that users are more likely to reuse products if positive emotions can be evoked.

Influence on user experience: Interface design also influences users' reflective evaluation of their experience. Norman [16] describes the reflective level as the stage at which users consciously interpret and evaluate their interaction with a system. This level involves higher-level cognitive processes, including interpretation, judgment, and subjective assessment of usability and satisfaction, but also meaning of a system or its use. The reflective level is influenced by the two other levels.

Interfaces that are visually organized according to Gestalt laws may enhance user experience by facilitating efficient interaction and eliciting positive emotional responses. Norman [16] argues that the overall impact of a product comes from reflection. Thus, positive visceral and behavioral experiences can contribute to more favorable reflective evaluations of the user experience of the system. When users can easily understand and interact with an interface and experience positive emotional responses during interaction, they are more likely to evaluate the system positively. In contrast, interfaces that increase cognitive effort or evoke negative emotional responses may lead to poorer overall user experience.

Taken together, Norman's emotional design framework provides a theoretical foundation for understanding how Gestalt laws influence human-computer interaction across multiple levels of processing. By shaping perceptual organization at the visual level, influencing emotional responses at the visceral level and ultimately affecting reflective evaluation, Gestalt laws play a central role in determining users' overall experience with digital interfaces.

Based on this theoretical foundation, this study examines the effects of the two Gestalt laws proximity (near vs. far) and similarity (similar vs. non-similar) on multiple outcome measures related to the three different levels. As shown in Fig. 12.

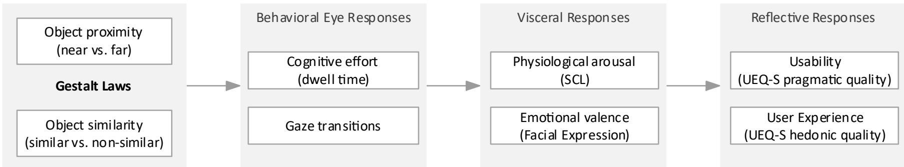


Fig. 12: Research model

Methodology

To answer the research questions, an experiment will be conducted in an e-commerce context.

In the present study, we focus on the Gestalt laws proximity and similarity, as both are well established in relation to usability, user interface, and user experience (number of ACM publications[5]: proximity = 457, similarity = 456). These two laws are among the most fundamental and frequently applied principles in visual interface design, as they directly influence how users perceptually group and interpret interface elements [12]. Moreover, proximity and similarity can be systematically manipulated in controlled experimental settings while maintaining functional equivalence of interfaces, making them especially suitable for isolating their causal effects on outcomes.

By using eye tracking, we primarily focus on dwell time and number of gaze transitions. We argue that stronger application of both Gestalt laws will reduce visual exploration and cognitive effort.

Skin conductance level (SCL) enables the assessment of physiological arousal. While it allows the measurement of arousal intensity, it does not provide information about emotional valence; however facial analysis does and hence makes possible the identification of discrete emotions such as joy or anger. We argue that stronger application of the two Gestalt laws proximity and similarity will lead to more positive emotions as the user interface will be perceived as clearer and easier to understand.

The UEQ-S [20] will be used as a standardized measure of user experience, capturing the two dimensions of pragmatic and hedonic quality. We argue that the perceived user experience will be higher, the more the two Gestalt laws are applied.

[5] We performed a literature search via ACM Digital Library by using the following search string: AllField:("gestalt" AND ("*name of the Gestalt Law*")) AND AllField:("user interface" OR "usability" OR "user experience"). We got the following results: Color (727), Balance / Symmetry (544), Proximity (457), Similarity (456), Prägnanz / Good Form / Simplicity / Succintness (390), Continuation / Continuity (374), Unity / Harmony (263), Closure (260). (2026-04-07).

Participants

Twenty-four undergraduate students will participate in the experiment. They will be recruited from an Austrian university. The participants will take part voluntarily in the study. All participants will be familiar with digital devices and e-commerce platforms. Participants will rate their familiarity with the Gestalt laws proximity and similarity on a scale from 1 (unfamiliar) to 7 (familiar).

Experimental Design

For the study, a 2 (similarity: similar vs. non-similar) x 2 (proximity: near vs. far) within-subject experimental design will be used. Therefore, each subject will participate in all four conditions. To avoid any learning or fatigue effects and to account for possible carry-over effects, a balanced latin-square design will counter-balance the conditions between the participants. The study will employ four experimental conditions that systematically vary the spatial arrangement (proximity: near vs. far) and the visual differentiation of the product card information elements (product name, product description, color, rating, price) underneath the product image in each product card (similarity: similar vs. non-similar).

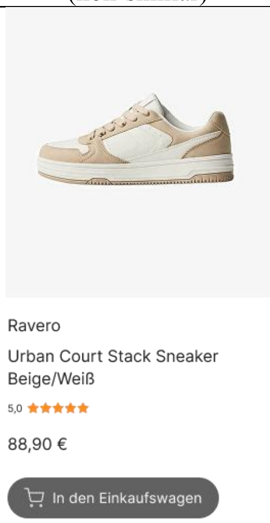
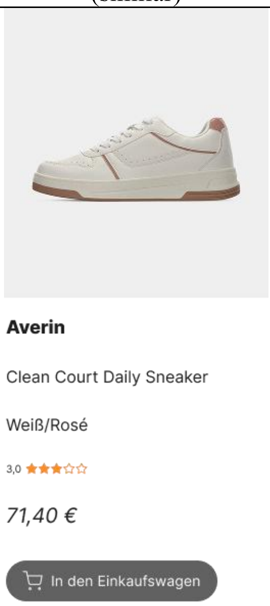
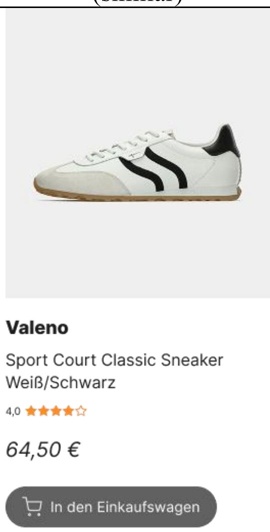
In the far non-similar condition (A), the product card information elements will be evenly distributed (same distance between the different labels) between two product images, and the text style is for all elements the same. In the near non-similar condition (B), the product card information elements will be grouped together underneath the related product image, and the text style is for all elements the same. In the far similar condition (C), the product card information elements will be evenly distributed between two product images and use the same text style is used for each category of information elements (e.g. same text style for prices). Finally, in the near similar condition (D), the product card information elements will be grouped together underneath the related product image, and the same text style is applied for each category of information elements. Table 8 illustrates the product card in the four different conditions.

For each condition, participants will perform the same search task. The participant will be instructed to search for a product with specific attributes. To find the right product, the participant will need to check the whole product card, especially the information underneath. The position of the product to be identified will vary between the conditions. A sample for the whole user interface is shown in Figure 13.

Procedure

Each session will start with signing a consent form obtained from each participant. A pre-test questionnaire will be given in advance to track some demographic data (age, gender), technical and domain knowledge (experience with e-commerce systems; familiarity with Gestalt laws). Next, a calibration of the eye-tracking equipment will be performed. Afterwards, each participant will complete each of the four conditions. Within each experiment, subjects will be provided with instructions and then presented with experimental conditions (visual stimuli) and will be instructed to complete the search task. After each condition, the UEQ-S [20] will be used to measure the perceived usability and user experience of the interface.

Table 8: Product card in the four different conditions.

<table>
<tr><th>Stimuli A: Proximity (far), Similarity (non-similar)</th><th>Stimuli B: Proximity (near), Similarity (non-similar)</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><th>Stimuli C: Proximity (far), Similarity (similar)</th><th>Stimuli D: Proximity (near), Similarity (similar)</th></tr>
<tr><td>
</td><td>
</td></tr>
</table>

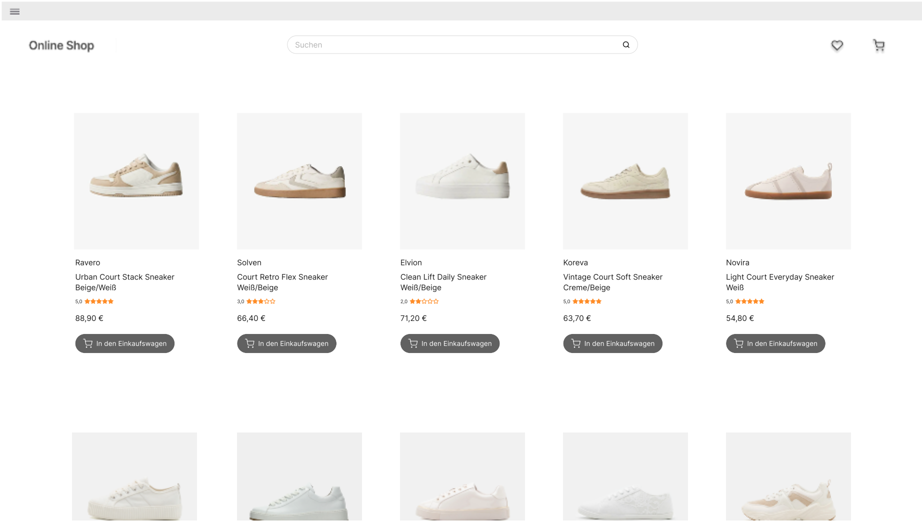


Figure 13: Search task user interface (proximity - near, similarity - non-similar).

Data Collection

Data collection will be conducted using the iMotions biometric platform including the collection of eye-tracking data with the SmartEye Aurora sensor (250 Hz) and galvanic skin response with the Shimmer3 iMotions GSR Kit. The GSR sensor will monitor changes in the skin conductivity between two electrodes attached to two fingers of one hand. For the facial coding the iMotions platform uses the automated facial coding AI Affectiva's AFFDEX [4] and is based on the facial action coding system FACS [9]. The following emotions and their intensity will be recognized by the mentioned system: Anger, Sadness, Disgust, Fear, Joy, Surprise and Contempt. Besides that, head movements, blinks, and overall valence can be measured.

Image presentations will be executed on a 1920 x 1020 resolution monitor. Initial analysis of the collected data will be performed using the iMotions software (v11).

Current Limitations

Choice and number of study participants: Participants are recruited from a single local university and share a similar field of study and age range. A more diverse sample (e.g. varying professional background, different age groups) combined with larger sample size would strengthen the findings. The current sample size allows for the identification of tendencies and hence is exploratory in nature.

Physiological measures: With SCL measurement and automated facial expression analysis the study uses two different physiological measures, which are both sensitive to noise, including body movements or inter-individual variability in physiological responsiveness [1]. To overcome the inter-individual differences, data will be normalized as a pre-processing step.

Selection of two Gestalt laws: The study focused on two Gestalt laws. While this allows controlled investigation, the interaction effects of other Gestalt laws remain unexplored. Future work should examine combinations of several laws to better capture their influence on visceral, behavioral, and reflective responses.

Selection of application context: Gestalt laws are applied across visual, auditory and haptic input and output in diverse contexts. This study focuses exclusively on visual output in an e-commerce context, which constrains validity for other contexts. Although the findings are expected to be broadly informative, validation in other domains and across different input and output modalities is necessary.

Conclusion

In summary, in this research-in-progress study we describe a lab experiment in the e-commerce context investigating the visceral, behavioral, and reflective responses to the two Gestalt laws proximity and similarity. We expect that variations in stimulus design will lead to differences in physiological responses, eye behavior, and self-reports. Our research findings can contribute to the design of more effective and enjoyable user interfaces because, by building on them, we can expect that the physical processes occurring in the user, visual perception, and ultimately conscious perception – all of which relate to the user experience – will be improved if the design principles examined here are correctly applied during the design process.

References

1. Ahmad, Z., Khan, N.: A Survey on Physiological Signal-Based Emotion Recognition. Bioengineering (Basel, Switzerland) **9**(11) (2022). doi: 10.3390/bioengineering9110688
2. Azmy, A., qutb, M., Nada, M.: Eye Tracking in Outdoor Advertising: Exploring the Application of Gestalt Principles. International Design Journal **13**(6), 437–446 (2023)
3. Bačić, D.: Exploring the Impact of Gestalt Laws of Human Perception in Business Information Visualization Context: An Eye-Tracking Study. In: 2024 47th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, pp. 1361–1366. IEEE (2024)
5. Bonnardel, N., Piolat, A., Le Bigot, L.: The impact of colour on Website appeal and users' cognitive processes. Displays **32**(2), 69–80 (2011)
6. Burch, M., Wallner, G., Ah, I. von, Pehar, F.: A User Evaluation on Mouse Movements in Visual Interfaces. In: Wallner, G., She, J., Burch, M., Liang, H.-N. (eds.) Proceedings of the 18th International Symposium on Visual Information Communication and Interaction. VINCI 2025: Proceedings of the 18th International Symposium on Visual Information Communication and Interaction, Linz Austria, 01 12 2025 03 12 2025, pp. 1–8. ACM, New York, NY, USA (2025). doi: 10.1145/3769534.3769536
7. Chang, D., Dooley, L.S., Tuovinen, J.E.: Gestalt theory in visual screen design: a new look at an old subject. In: McDougall, A., Murname, J., Chambers, D.

(eds.) CRPIT '02: Proceedings of the Seventh world conference on computers in education conference on Computers in education: Australian topics, Copenhagen, Denmark, pp. 5–12. Australian Computer Society, Inc, AUS (2002)
8. Ehrenfels, C. von: Über Gestaltungsqualitäten. Vierteljahresschrift für wissenschaftliche Philosophie, 249–292 (1890)
9. Ekman, P., Friesen, W.V., Hager, J.C.: Facial Action Coding System, Bd. 2. Human Face (2002)
10. Hartson, H.R., Pyla, P.S.: The UX book. Agile UX design for a quality user experience. Morgan Kaufmann, Cambridge MA (2019)
11. Juanéda, C., Sénécal, S., Léger, P.-M.: Product Web Page Design: A Psychophysiological Investigation of the Influence of Product Similarity, Visual Proximity on Attention and Performance. In: Nah, F.F.-H., Xiao, B.S. (eds.) HCI in Business, Government, and Organizations, pp. 327–337. Springer International Publishing, Cham (2018)
12. Lidwell, W.: Universal principles of design : 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design. Rockport Publishers, Beverly, MA (2010)
13. Ma, C., Wang, H., Wu, J., Xue, C.: Applying gestalt similarity to improve visual perception of interface color quantity: An EEG study. International Journal of Industrial Ergonomics **100**, 103521 (2024)
14. Möller, B., Brezing, C., Unz, D.: What should a corporate website look like? The influence of Gestalt principles and visualisation in website design on the degree of acceptance and recommendation. Behaviour & Information Technology **31**(7), 739–751 (2012)
15. Nissen, A., Riedl, R., Schütte, R.: Users' reactions to website designs: A neuroimaging study based on evolutionary psychology with a focus on color and button shape. Computers in Human Behavior **155**, 108168 (2024)
16. Norman, D.A.: Emotional design: why we love (or hate) everyday things. Basic Books, Inc, New York, NY, USA (2004)
17. Portnoff, S.R.: The introductory computer programming course is first and foremost a language course. ACM Inroads **9**(2), 34–52 (2018). doi: 10.1145/3152433
18. Promann, M., Brunswicker, S.: The Effect of Proximity in Social Data Charts on Perceived Unity. 2018 IEEE Conference on Visual Analytics Science and Technology, VAST 2018 - Proceedings (2018). doi: 10.1109/VAST.2018.8802449
19. Ripalda, D., Guevara, C., Garrido, A.: Framework based on gestalt principles to design mobile interfaces for a better user experience. Advances in Intelligent Systems and Computing **1217 AISC**, 158–165 (2020). doi: 10.1007/978-3-030-51828-8_21
20. Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). IJIMAI **4**(6), 103 (2017). doi: 10.9781/ijimai.2017.09.001

Autistic Traits, Executive Functions, and Burnout in IT Professionals

Ronnie Jia[1] and Yulia Sullivan[2]

[1] University of North Texas, Denton, Texas, USA
RonnieJia@Gmail.com
[2] Baylor University, Waco, Texas, USA
Yulia_Sullivan@baylor.edu

Abstract. Employee burnout is one of the main challenges for organizations. It has been shown to cause long-term psychological and physiological consequences for the employee. Burnout is particularly prevalent among IT workers, given the characteristics of IT professions. Although information systems (IS) research has predominantly explained burnout among IT workers through job and workplace characteristics, comparatively less attention has been given to individual differences that may shape vulnerability. Building on emerging evidence of elevated autistic traits among IT professionals, this study investigates whether higher autistic traits among IT employees are associated with weaker executive functions, which in turn predicts their susceptibility to burnout. By shifting attention from workplace drivers alone to underlying individual-level mechanisms, this work extends IS burnout research and offers actionable implications for interventions aimed at improving IT employee well-being, retention, and productivity.

Keywords: Autistic traits; executive functions; burnout; IT professionals

Introduction

IT professionals are widely regarded as a strategic asset that enables organizations to enhance their competitive advantage [1-2]. However, despite their critical role in supporting organizational strategy [3], the IT sector is often characterized as a high- pressure work environment in which employees must keep pace with rapidly evolving technical skills and demands [2]. As a result, IT employee burnout has been a central concern in IS literature [3]. Much of this research is in keeping with Job-Demand Resources (JD-R) framework [4-5], which conceptualizes burnout as arising when job demands (i.e., aspects of the job that require sustained physical or mental effort) exceed available job resources (i.e., physical, psychological, social, or organizational aspects of work that help achieve goals). Within this tradition, research has demonstrated how job-related stressors, such as workload and time pressure, contribute to burnout among IT professionals [3]. However, due to the emphasis on workplace and job characteristics, less attention has been paid to employees' internal cognitive resources and constraints.

This omission may be particularly consequential in the IT context, given growing evidence that IT professionals might exhibit distinctive cognitive and psychological profiles.

Recent empirical studies show that autistic traits are linked to one's intrinsic interest in IT [6] and that IT professionals, relative to the general population, tend to exhibit elevated levels of autistic traits [7]. Importantly, this work conceptualizes autistic traits as continuous individual differences rather than diagnostic categories [6-7]. At the same time, IT professionals report disproportionately high levels of anxiety and depression symptoms [7], raising questions about why a field that strongly attracts individuals with autistic traits may also expose them to heightened mental health risk. This pattern points to an apparent double-edged dynamic: autistic traits may contribute to strong person–job fit in technical work, while simultaneously increasing vulnerability to burnout.

Further, existing IS studies have primarily focused on the challenges and opportunities faced by autistic IT professionals (e.g., [8-10]), as well as impacts of neurodiverse traits in technology users (e.g., [11]). However, this line of research remains largely descriptive and outcome-oriented, offering limited insight into the underlying cognitive mechanisms that shape how autistic individuals engage with and respond to work demands in IT settings. This line of research remains largely descriptive and outcome-oriented, offering limited insight into the underlying cognitive mechanisms that shape how autistic individuals engage with and respond to work demands in IT settings.

To address these gaps, we adopt a trait-based perspective that extends the JD-R framework by incorporating internal cognitive resources and constraints. In particular, we focus on executive functions—higher-order cognitive processes involved in regulating attention, working memory, flexibility, and inhibition—as a key mechanism through which autistic traits may influence burnout risk. Accordingly, our research is guided by two questions: 1) Do higher levels of autistic traits among IT professionals predict lower executive functions? and 2) Do lower executive functions contribute to a greater tendency toward burnout?

To answer these questions, we conduct a study of IT professionals that measures autistic traits and assesses executive functions using computer-based, game-like tasks grounded in established neuropsychological methods. By examining burnout through the joint lenses of autistic traits and executive functions, this research extends the JD-R framework and advances a more cognitively informed understanding of stress and wellbeing in the IT workforce.

Theory Development

Autism and Autistic Traits

As specified in *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition [12], autism is a neurodevelopmental condition characterized by (a) pervasive deficits in reciprocal social communication and interaction (e.g., challenges in initiating and sustaining conversations and developing and maintaining friendships), and (b) restricted, repetitive patterns of behavior, interests, or activities (e.g., stereotyped movements, insistence on routines). Despite its categorical diagnosis status (positive/negative), autism is increasingly understood as reflecting a constellation of autistic traits that are continuously distributed throughout the general population [6-7]. In view of the "normal distribution of traits, rather than a bimodal distribution," "it does appear that… one can be 'a bit autistic'" ([13], p. 223) as most individuals only exhibit subclinical levels of autistic traits.

Although autism has long been characterized in the popular press as an "engineer's disorder" [14] and an "open secret" within the IT profession [15], systematic examinations of the relationship between autism and IT have only recently emerged in the IS field [6-7]. This research suggests that traits associated with systemizing, analytical reasoning, and technical focus may draw individuals toward IT careers [6], while simultaneously shaping how they experience workplace social and emotional demands [7].

To understand how autistic traits translate into differential vulnerability to burnout, however, it is necessary to identify the underlying cognitive mechanisms through which these traits affect demand management. One such mechanism is executive functioning.

Executive Functions

Executive functions refer to a set of higher order cognitive processes that regulate goal directed behavior and adaptive functioning [16]. These processes rely on distributed neural networks involving the prefrontal cortex, basal ganglia, and thalamus [17]. Executive functions are typically conceptualized as comprising three core components: working memory (i.e., the ability to maintain and update task relevant information), cognitive flexibility (i.e., the ability to shift between tasks or mental sets), and inhibitory control (i.e., the ability to suppress dominant but task irrelevant responses) [18].

According to executive function theory of autism, autistic individuals exhibit impairments in mental control abilities necessary for maintaining problem solving strategies in the service of future goals [19]. Empirical studies consistently show that autistic individuals tend to perform worse on tasks requiring cognitive flexibility, higher working memory load, and spatial working memory (e.g., [20]). Consistent with these findings, deficits in core executive subfunctions—particularly working memory and cognitive flexibility—have been linked to key autism related outcomes, including repetitive behaviors in adults [19] [21]. Evidence regarding inhibitory control is comparatively mixed, suggesting a domain specific pattern of executive dysfunction rather than a generalized deficit.

Research on working memory in autism has produced heterogeneous findings. One study [22] reported no significant autism-related differences in working memory, whereas a meta-analysis [23] concluded that working memory is, on average, impaired in autism, with spatial working memory more strongly affected than verbal working memory. Kercood et al. [20] similarly reported working memory impairments among autistic individuals. Taken together, this body of evidence suggests that executive function differences in autism are most pronounced in flexibility and working memory processes—capacities that are especially relevant in cognitively demanding technical work environments.

Research Model

Figure 1 illustrates the proposed research model. Burnout, defined as the state of energy depletion and disengagement from one's work [24], has been consistently linked to cognitive impairment [24-26]. Prior research suggests a bidirectional relationship between burnout and cognitive functioning. On one hand, executive dysfunction may increase vulnerability to burnout when workplace demands exceed one's cognitive capacity; on the other hand, burnout can undermine memory and other cognitive functions [25]. Given this potential reciprocal relationship, impaired executive functions may heighten burnout risk, while prolonged burnout may further erode executive functioning over time. In the present study, we focus on this relationship within a cross-sectional design to establish baseline associations.

While prior research indicates that IT professionals, on average, exhibit higher autistic traits and vulnerability to mental health challenges than the general population [6-7], the present study focuses on two hypotheses that extend this literature by specifying an underlying neurocognitive mechanism:

H1: Higher autistic traits are associated with lower executive functions among IT professionals.

H2: Lower executive functions are associated with higher burnout among IT professionals.

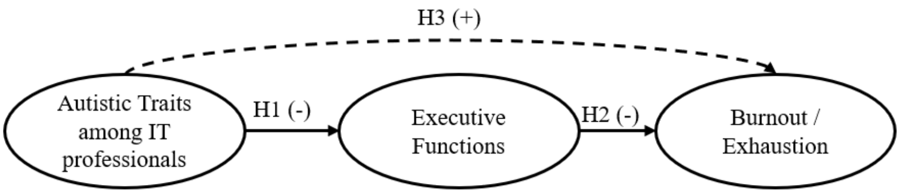


Fig. 14. Research Model

Together, these hypotheses imply a mediating role of executive functions in the relationship between autistic traits and burnout. Supporting this view, prior studies show that individuals with higher autistic traits perform worse on classic executive function tasks, such as the Wisconsin Card Sorting Test and the Tower of Hanoi [27], indicating impairments in cognitive flexibility and behavioral regulation [20].

Building on the JD-R framework, we conceptualize executive functions as a critical personal resource that shapes individuals' capacity to cope with job demands. From this perspective, executive dysfunction constrains the ability to allocate, sustain, and replenish cognitive resources, thereby increasing burnout risk. This vulnerability may be particularly pronounced in IT contexts, where work demands extend beyond technical problem solving to include navigating social interactions, conceptualizing abstract requirements, organizing tasks, maintaining punctuality, and adapting to unfamiliar or rapidly changing situations [28]. Over time, the cumulative strain associated with managing these demands under constrained executive functioning may contribute to heightened exhaustion and disengagement. Thus, our third hypothesis is

H3: Higher autistic traits are associated with higher burnout through reduced executive functions

As discussed earlier, our cross-sectional design does not allow causal inference or establish temporal ordering of the research model. We thus do not infer causal direction for the two relationships—neither between autistic traits and executive functions, nor between executive functions and burnout.

Methodology

IT professionals' autistic traits were assessed using the 50-item Autism-Spectrum Quotient (AQ) [29]. The AQ yields scores ranging from 0 to 50 and provides normative data as well as a recommended clinical cut point; scores of 26 or higher have been suggested as indicating that an individual may meet a diagnostic threshold. Using this cut point, prior work reports an overall classification accuracy of 83% [30]. As noted earlier, however, the AQ is not a diagnostic instrument. We used Maslach burnout inventory, with the focus on exhaustion, to measure burnout.

To assess executive functions, we followed the task protocol in Sullivan et al. [18] and used neuropsychological tasks (e.g., N-back, color-shape) to measure executive functions (Table 1). Although these tasks do not directly examine neural activity in the way neuroimaging methods do, they are well-established neuropsychological paradigms for assessing cognitive processes linked to brain function [31].

Table 9. Executive Functions Measures [18]

Task	EF	Explanation
N-back	Working memory	Participants must continuously update working memory by keeping track of the most recent n stimuli in a sequence.
Color-shape	Cognitive Flexibility	Depending on the stimulus cue, participants switch between making color-based judgments and shape-based judgments.
Stroop	Inhibition	Participants must inhibit an automatic response (e.g., reading the word) in order to produce the task-relevant response (e.g., naming the ink color).

This study has received Institutional Review Board (IRB) approval, and data collection is currently ongoing. A sample of IT professionals is being recruited via Prolific to complete a two-part study. Prolific is an online research platform that enables access to diverse and targeted participant pools. It also allows for prescreening based on occupational criteria, which facilitated the recruitment of participants in IT occupations. Participants first completed survey scales assessing their autistic traits and burnout, and then performed the neuropsychological tasks assessing executive functions administered through the Millisecond Inquisit online platform.

Of the 184 participants so far, 169 met attention checks and completed the AQ. After excluding those who did not complete the working memory (n-back) task, the current useful sample consisted of 150 participants.

Preliminary Results

Autistic Traits

Of the 169 IT professionals who completed the AQ scale, 24.3% met the cut-off ($\geq$ 26), indicating elevated autistic traits consistent with probable autism based on this screening threshold. This proportion is considerably higher than the prevalence in the U.S. general population (3.05%) [32]. This result underscores why our research questions matter: a considerable proportion of the IT workforce exhibits clinically significant levels of autistic traits, making it important to examine whether their executive functions differ and whether such differences relate to burnout risk. In our analyses, we control for demographic and professional characteristics, such as gender, age, and IT job role.

Executive Functions and Employee Burnout

In preliminary analyses, we also observed a statistically significant negative correlation between working memory performance and exhaustion ($p < 0.001$), the core dimension of burnout, such that lower working memory was associated with higher exhaustion. Although we have not yet conducted the full set of planned analyses, such as multivariate models and robustness checks, these initial results are consistent with our research questions and suggest that executive functions, particularly working memory, may be meaningfully related to burnout risk among IT professionals.

Expected Contributions

Going beyond job demands and resources, this paper proposes an executive function perspective on IT employee burnout. If systematic differences in executive functions do exist among IT professionals, what does this mean for how organizations manage and support technical employees? How can this knowledge be used to improve IT employee wellbeing, retention, and productivity? By integrating neurocognitive differences into burnout research, this work can inform more targeted interventions and contribute to a healthier, happier, and more productive IT workforce.

Further, our research highlights the role of individual differences in shaping workplace experiences, particularly among IT professionals. In this regard, prior research on individual characteristics related to burnout has often been examined through the lens of technostress, showing that technology-mediated interruptions can increase stress and impair task performance [33-35]. However, these effects may not be uniform across individuals. In particular, autistic IT professionals may bring distinct cognitive strengths to their work. For example, they are often drawn to IT roles due to strong intrinsic interest [28], which may support sustained engagement and performance in technology-intensive environments.

At the same time, some evidence suggests that certain neurodivergent traits are associated with strengths in memory and information processing, which could facilitate the retention and management of large amounts of information. These strengths may, in some cases, buffer against the negative effects of technostressors that are typically observed among non-IT or neurotypical populations. Taken together, this perspective underscores the importance of accounting for neurodivergent profiles when examining how technostress influences employee outcomes.

Looking ahead, future research should further explore the potential benefits associated with autism in IT work, particularly in relation to resilience against technostress and burnout. Such work could help move the literature beyond deficit-based perspectives toward a more balanced understanding of how neurodiversity might contribute to the IT workplace. In line with this direction, our ongoing research will incorporate eye-tracking methods to capture fine-grained indicators of executive function processes among IT students, enabling us to examine attentional allocation, cognitive load, and information processing strategies during task performance.

References

1. Ferratt, T.W., Prasad, J., Enns, H.G.: Synergy and its Limits in Managing Information Technology Professionals. Inf. Syst. Res. 23(4), 1175–1194 (2012)
2. Zaza, S., Joseph, D., Armstrong, D.J.: Are IT Professionals Unique? A Second Order Meta Analytic Comparison of Turnover Intentions Across Occupations. MIS Q. 47(3), 1213–1238 (2023)
3. Moore, J.E.: One Road to Turnover: An Examination of Work Exhaustion in Technology Professionals. MIS Q. 24(1), 141–168 (2000)
4. Bakker, A. B., Demerouti, E.: Job Demands–Resources Theory: Taking Stock and Looking Forward. J. Occup. Health Psychol. 22(3), 273–285 (2017)
5. Demerouti, E., Bakker, A.B., Nachreiner, F., Schaufeli, W.B.: The Job Demands–Resources Model of Burnout. J. Appl. Psychol. 86(3), 499–512 (2001)
6. Jia, R., Steelman, Z.R., Jia, H.H.: What Makes One Intrinsically Interested in IT? An Exploratory Study on Influences of Autistic Tendency and Gender in the U.S. and India. MIS Q. 46(3), 1603–1634 (2022)
7. Jia, R., Riemenschneider, C., Jia, H.H.: Autistic Traits, Theory of Mind, and Implications for IT Professionals' Emotional Labor and Mental Wellbeing. MIS Q. 50 (2026)
8. Holmes, T., Annabi, H.: The Dark Side of Software Development: Job Stress Amongst Autistic Software Developers. In: Proc. 53rd Hawaii Int. Conf. Syst. Sci. (HICSS), pp. 1–10 (2020)
9. Loiacono, E.T., Ren, H.: Building a Neurodiverse High-Tech Workforce. MIS Q. Exec. 17(4), 1540–1960 (2018)
10. Sundermeier, J., Birkner, S., Ettl, K., Kensbock, J., Tegtmeier, S.: Hello Diversity! Opportunities and Challenges of Entrepreneurial Diversity in the Digital Age. Commun. Assoc. Inf. Syst. 47, 697–715 (2020)
11. Turel, O., Bechara, A.: Time to Talk About It: Neurodiversity of Information Systems Users in Leisure Settings. J. Assoc. Inf. Syst. 25(5), 1343–1371 (2024)

12. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (5th ed., text rev.; DSM 5 TR). American Psychiatric Association Publishing (2022)
13. Happé, F., Frith, U.: Annual Research Review: Looking Back to Look Forward—Changes in the Concept of Autism and Implications for Future Research. J. Child Psychol. Psychiatry 61(3), 218–232 (2020)
14. Silberman, S.: The Geek Syndrome. Wired (2001). https://www.wired.com/2001/12/aspergers/
15. Mayor, T.: IT's Open Secret. ComputerWorld, 23–27 (2008, May 5)
16. Miyake, A., Friedman, N.P.: The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. Curr. Dir. Psychol. Sci. 21(1), 8–14 (2012)
17. Sinzig, J., Morsch, D., Bruning, N., Schmidt, M.H., Lehmkuhl, G.: Inhibition, Flexibility, Working Memory and Planning in Autism Spectrum Disorders with and without Comorbid ADHD Symptoms. Child Adolesc. Psychiatry Ment. Health 2(1) (2008). doi:10.1186/1753-2000-2-4
18. Sullivan, Y.W., Davis, F.D., Koh, C.: Executive Functions and Information Systems Learning. MIS Q. 46(2), 813–880 (2022)
19. Lopez, B.R., Lincoln, A.J., Ozonoff, S., Lai, Z.: Examining the Relationship between Executive Functions and Restricted, Repetitive Symptoms of Autistic Disorder. J. Autism Dev. Disord. 35(4), 445–460 (2005)
20. Kercood, S., Grskovic, J.A., Banda, D., Begeske, J.: Working Memory and Autism: A Review of Literature. Res. Autism Spectr. Disord. 8(10), 1316–1332 (2014)
21. Turner, M.: Annotation: Repetitive Behaviour in Autism: A Review of Psychological Research. J. Child Psychol. Psychiatry Allied Discip. 40(6), 839–849 (1999)
22. Ozonoff, S., Strayer, D.L.: Further Evidence of Intact Working Memory in Autism. J. Autism Dev. Disord. 31(3), 257–263 (2001)
23. Wang, Y., Zhang, Y.B., Liu, L.L., Cui, J.F., Wang, J., Shum, D.H.K., Chan, R.C.K.: A Meta Analysis of Working Memory Impairments in Autism Spectrum Disorders. Neuropsychol. Rev. 27(1), 46–61 (2017)
24. Lemonaki, R., Xanthopoulou, D., Bardos, A.N., Karademas, E.C., Simos, P.G.: Burnout and Job Performance: A Two Wave Study on the Mediating Role of Employee Cognitive Functioning. Eur. J. Work Organ. Psychol. 30(5), 692–704 (2021)
25. Deligkaris, P., Panagopoulou, E., Montgomery, A.J., Masoura, E.: Job Burnout and Cognitive Functioning: A Systematic Review. Work Stress 28, 107–123 (2014)
26. Jonsdottir, I.H., Nordlund, A., Ellbin, S., Ljung, T., Glise, K., Währborg, P., Wallin, A.: Cognitive Impairment in Patients with Stress Related Exhaustion. Stress 16(2), 181–190 (2013)
27. Kleinhans, N., Akshoomoff, N., Delis, D.C.: Executive Functions in Autism and Asperger's Disorder: Flexibility, Fluency, and Inhibition. Dev. Neuropsychol. 27(3), 379–401 (2005)
28. Tomczak, M.T., Kulikowski, K.: Toward an Understanding of Occupational Burnout Among Employees with Autism—The Job Demands Resources Theory Perspective. Curr. Psychol. 43(2), 1582–1594 (2024)
29. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E.: The Autism Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High Functioning Autism, Males and Females, Scientists and Mathematicians. J. Autism Dev. Disord. 31(1), 5–17 (2001)
30. Woodbury Smith, M.R., Robinson, J., Wheelwright, S., Baron Cohen, S.: Screening Adults for Asperger Syndrome Using the AQ: A Preliminary Study of Its Diagnostic Validity in Clinical Practice. J. Autism Dev. Disord. 35(3), 331–335 (2005)

31. León-Domínguez, U., Martín-Rodríguez, J.F., León-Carrión, J.: Executive n back Tasks for the Neuropsychological Assessment of Working Memory. Behav. Brain Res. 292, 167–173 (2015)
32. Zablotsky, B., Ng, A.E., Black, L.I., Blumberg, S.J.: Diagnosed Developmental Disabilities in Children Aged 3–17 Years: United States, 2019–2021. NCHS Data Brief, No. 473. National Center for Health Statistics, Hyattsville, MD (2023)
33. Maier, C., Laumer, S., Wirth, J., Weitzel, T.: Technostress and the Hierarchical Levels of Personality: A Two Wave Study with Multiple Data Samples. Eur. J. Inf. Syst. 28(5), 496–522 (2019)
34. Riedl, R., VanMeter, K.: Explaining Inconsistent Research Findings on the Relationship Between Age and Technostress Perceptions: Insights from the Neuroscience Literature. In: NeuroIS Retreat, pp. 307–320. Springer, Cham (2020)
35. Tams, S., Thatcher, J.B., Grover, V.: Concentration, Competence, Confidence, and Capture: An Experimental Study of Age, Interruption Based Technostress, and Task Performance. J. Assoc. Inf. Syst. 19(9), 2 (2018)

Neurophysiological Correlates of Flow: A Systematic Review based on Csikszentmihalyi's Flow Theory

Cosima von Uechtritz[1], Mario Nadj[1]

[1] Rhine-Ruhr Institute of Information Systems, University of Duisburg-Essen, Germany
{cosima.uechtritz, mario.nadj}@ris.uni-due.de

Abstract. Csikszentmihalyi's flow theory is increasingly studied in NeuroIS to better understand user behavior and inform the design of flow-supportive systems. Prior reviews have focused on either the peripheral or central nervous system or on specific contexts. Moreover, research is complicated by measurement heterogeneity, as instruments often omit or modify the flow dimensions, and task heterogeneity, since physically active tasks elicit distinct physiological responses. Against this backdrop, we conducted a systematic and up-to-date literature review of neurophysiological flow correlates using Csikszentmihalyi's theory as ground truth and differentiating by task type. Our findings reveal persistent heterogeneity of neurophysiological activation patterns, shaped by the task, participant factors, and methodological choices. On this basis, we highlight critical pitfalls and outline directions for future research on the neurophysiological correlates of flow.
Keywords: Flow theory • Neurophysiology • Psychophysiology • Systematic review.

Introduction

Flow theory by Csikszentmihalyi [1] has gained considerable attention in information systems [2]. Flow describes the state of complete task absorption [3] and is related to improved well-being and peak performance [4]. This state is relevant, among others, in knowledge work, where employees are frequently interrupted by information technology. Developing a flow-supportive system that can automatically suppress task-irrelevant notifications could help employees maintain focus [5]. Achieving this, however, requires tools capable of continuously assessing the flow state. One promising approach involves measuring neurophysiological parameters, such as those captured through electrocardiogram chest belts [6] or portable electroencephalography (EEG) [7]. In addition, recent advances in sensor technology, such as webcam-based remote assessment of eye activity and heart rate, have expanded analytical and experimental possibilities in NeuroIS, supporting non-obstructive flow measurement.

These advancements underscore the need to integrate neurophysiological flow research findings. In the last decade, several systematic literature reviews (SLRs) have addressed aspects of this domain: [8] focuses on the peripheral nervous system; [9, 10] examine the central nervous system (CNS), and [11] considers the video games context solely. However, as research on the neurophysiological correlates of flow continues to expand, and several new studies have emerged in recent years [5, 12–16], a systematic and up-to-date synthesis of findings is needed. Methodological heterogeneity further complicates the field, making it seem to prevent firm conclusions [9].

First, a source of inconsistency lies in measurement heterogeneity. Existing studies rely on different flow measurement instruments, many of which either do not cover all five characteristics of Csikszentmihalyi's theory (e.g., [17]) or modify them (e.g., [18]). The absence of a common ground truth limits comparability across studies. Second, task heterogeneity introduces additional complexity. Physically active tasks (e.g., dancing [19]) and predominantly cognitive tasks (e.g., knowledge work [12]) are likely to elicit distinct physiological activation patterns, complicating cross-study comparisons if not differentiated. To date, neurophysiological flow research has not been systematically evaluated regarding these sources of heterogeneity.

To address these gaps, we conducted an SLR to address the following research question: *What is the current state of the art on neurophysiological correlates of flow measurements based on Csikszentmihalyi's flow theory?* In contrast to the prior SLRs [8–11], we restrict inclusion to studies that operationalize flow using a ground truth derived from Csikszentmihalyi's theory, and we differentiate findings by task type. Our results reveal a persistently heterogeneous pattern of neurophysiological activations across studies, even when applying a common theoretical ground truth and task-based differentiation. We relate these findings to existing theoretical accounts linking flow to neurophysiology, highlight critical pitfalls, and delineate directions for future research.

Theoretical Background

Flow is conceptualized by (i) three antecedents – (1) challenge-skill balance, (2) clear goals and (3) unambiguous feedback, (ii) five characteristics – (4) action-awareness merging, (5) sense of control, (6) loss of self-consciousness, (7) transformation of time, (8) concentration, and (iii) consequences like (9) autotelic experience [3, 20]. The three antecedents, in particular the challenge-skill balance, in which an individual's perceived challenge and skill are matched, are frequently used as a methodological framework to explore how flow manifests across the CNS, autonomic nervous system (ANS), and endocrine system [21–23].

The sympathetic nervous system (SNS) – "*fight-or-flight*" – and the parasympathetic nervous system (PSNS) – "*rest-or-digest*"– are two distinct systems of the ANS[6] [25]. Based on the "*modes of autonomic control*", six ANS activation patterns exist, which can be differentiated into three categories [26]. First, *coupled reciprocal activation*, which indicates increasing activity in one branch related to decreasing activity in the other branch, leading to (i) reciprocal SNS or (ii) reciprocal PSNS activation patterns. Second, *coupled non-reciprocal activation* where both branches are positively correlated with each other: (iii) co-inhibition or (iv) co-activation. Third, *non-coupled*

---

[6] The SNS is activated in stressful situations and elicits reactions such as pupil dilation to prepare the body for a *"fight-or-flight"* response. In contrast, the PSNS is responsible for the *"rest-or-digest"* response in the body and leads to pupil contraction [24]. The enteric nervous system is beyond the scope of the present study and is therefore not discussed here.

activation where the activation of both branches is unrelated to each other (v) uncoupled PSNS or (vi) uncoupled SNS. As highlighted by opposing views (e.g., [27, 28]), the underlying relationship between flow and ANS activity does not share a common perspective among researchers.

Similarly, different theories exist regarding flow and the CNS. The *Transient Hypofrontality Hypothesis* (THH) explains the flow state based on the distinction between the explicit (connected to conscious awareness) and implicit (inaccessible to conscious awareness) systems [29]. A requirement for the flow state is the temporary suppression of the explicit system, which has been criticized as an oversimplification of the flow experience [30]. Thus, the *Synchronization theory of flow* addresses this issue by proposing that flow arises from the synchronization of attentional and reward networks, in a state of balance between challenge and skill. This synchronization is a complex system with energetic optimization, which leads to the pleasurable experience of effortless attention [30]. A third approach, the *internal model of flow*, has emerged with the aim of integrating both theories [31]. In this view, the brain uses an internal model that provides a sensorimotor representation of oneself within the surrounding environment. The flow state is thought to depend on these internal models, which are formed in the cerebellum during skill acquisition and are associated with the experience of intuitive and effortless behavior [31].

With respect to the endocrine system, comparatively little research has examined its relationship to the flow experience. The endocrine system regulates the production of hormones that influence physiological stress responses [32]. Research on flow has so far focused on the relation between cortisol and flow [33]. Evidence regarding the cortisol-flow relationship remains mixed, with some studies suggesting an inverted U-shaped pattern [23, 34] and others finding no significant association [35].

Method

We conducted an SLR following [36] with no date restriction in order to provide a comprehensive overview of the field. We applied the search term across four databases, resulting in 1,535 initial results. Following the removal of duplicates, 1170 records were retained for further screening.

First, the selection criteria (see Fig.1) were applied to the title, abstract, and keyword sections of the identified articles, excluding 1,093 articles. Next, the full texts of the remaining 77 articles were screened, resulting in 26 included articles. For instance, [17] was excluded because the subjective self-report to assess flow consists of three items *("I would love to solve math calculations of that kind again", "Task demands were well matched to my ability", and "I was thrilled"*), which do not cover the five characteristics of Csikszentmihalyi's flow theory (exclusion reason 3). Finally, a forward and backward search identified three additional articles. In total, the review included 29 relevant articles (please note that the 29 articles included 30 studies). We analyzed the studies through the lens of (a) flow and its dimensions, as well as (b) task type. In particular, studies examined flow either at the level of individual dimensions (see Chapter 2, first paragraph), as second-order constructs that group these dimensions into broader factors, such as *fluency*, reflecting a smooth and automatized process, and *absorption*, reflecting total immersion in the activity, or as *global flow* calculated as the average across all dimensions. As flow may be highly task-dependent [11], we derived task categories from recurring contexts in which flow has been empirically examined (e.g., gaming [11], sport [37], music [38]).

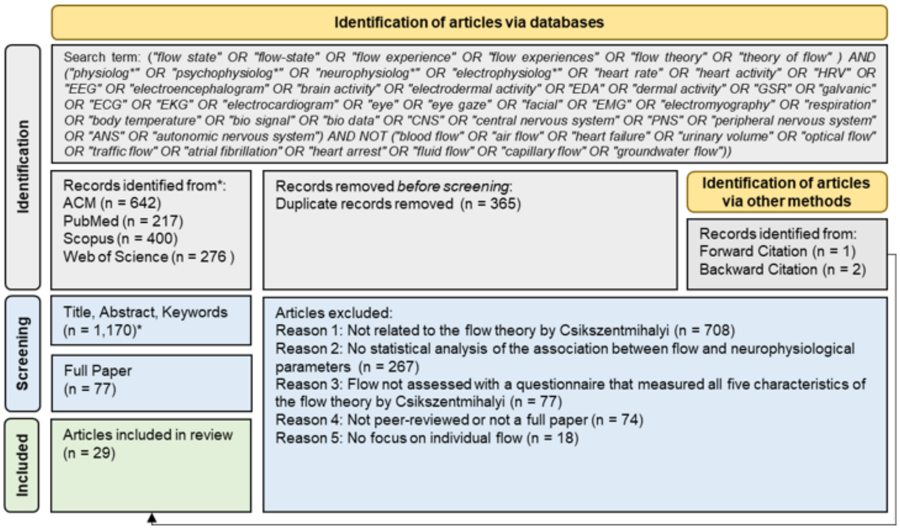


Fig 15: Description of the SLR process (* = two sources could not be retrieved)

Results

Due to space restrictions, the concept matrix underlying this review is available in an online appendix (Click here[7]). The following sections present findings across the 30 included studies, organized by study setup, ANS, CNS, and additional parameters, with the latter three further examined by flow dimension and task.

Study Setup

Task. Most studies investigated the flow experience in a gaming-related task (16 studies) using games, such as Pacman [34], Tetris [28, 39], or a VR Horror game [16]. Three studies investigated flow during music performances among pianists [40, 41] or mixed musician samples [13], three studies during sport, including running [42], cycling [43], and free throws [44], and three in a knowledge work context, including thesis writing [45], working on one's own project [12], and language learning [46]. The remaining studies focused on diverse tasks, such as dancing [19], filling in different questionnaires [47], or the Stroop stress test [48].

[7] https://osf.io/gv5pf/overview?view_only=050ae4eb149146c48ed175164aff3db7

Participants. Across the identified studies, the majority (19 studies) did not specify participant inclusion and exclusion criteria. Of the eleven remaining studies that specified participant inclusion criteria, definitions of expertise varied considerably. For example, pianists were classified as experts if they were soloists or students at the Royal Music Academy in Stockholm [41], or if they met the minimum standard for music program applicants in higher education [40]; dancers, if they had been financially remunerated for at least one project [19]; and gamers, if they had accumulated more than five hours of gameplay in the week prior to the experiment [49]. Four of the eleven studies that specify participant selection criteria included both experts and non-experts to examine group differences [13, 21, 49, 50].

Autonomous Nervous System

Global Flow. A total of seven studies investigated the relationship between flow and physiological activation using gaming tasks. Three studies suggest a co-activation pattern, [33, 39, 49], whereas two findings [22, 28] are most consistent with an uncoupled PSNS activation. No clear activation pattern could be identified in two studies [16, 47].

Furthermore, two studies used a piano-playing music task to examine flow [40, 41], concluding that flow is characterized by a co-activation pattern. One study used a Stroop stress test, indicating co-activation with predominant SNS activation [48], while another study focusing on filling in questionnaires does not indicate a clear activation pattern [47]. A co-inhibition has been solely identified in a dancing performance [19].

Absorption and Fluency (see Chapter 3)**.** Two studies [34, 49] investigated their findings regarding second-order constructs. Both studies suggest that absorption may represent a stronger indicator of flow in a gaming task. [34] found that the relationships between physiological markers and flow were primarily driven by absorption, relating to a co-activation pattern. Similarly, [49] demonstrated that the heart-evoked potential showed no correlation with global flow, but a positive relationship with absorption, whereas the autonomous activation patterns were only interpreted at the level of global flow.

Flow Dimensions (see Chapter 2, first paragraph)**.** Finally, two studies particularly examined flow dimensions, such as challenge skill (e.g., flow antecedent), sense of control (e.g., flow characteristic), or autotelic experience (e.g., flow consequence) in dancing [19] or gaming [51] (see Table 1 in the online appendix). However, no clear activation pattern can be identified among these studies.

Central nervous system

Global flow. A total of six studies focus on gaming, where three employed EEG measurements to assess brain activity across different spatial regions and frequency bands [14, 15, 49], whereas the remaining two studies used (functional) near-infrared spectroscopy (fNIRS) to compute oxygenated hemoglobin [21, 39] or the average oxygenation change [28]. Across these studies, two provide support for the THH [39, 49], and one study suggests support for the Synchronization Theory [21]. The findings of [14, 15, 28] do not provide a clear trend towards one of the three propositions.

Two studies examined flow during sport, including running [42] and cycling [43]. Single studies investigated flow music performance [13], and the observation of table tennis serves [50]. One study assessed two different tasks, specifically self-directed project work and a math task [12]. Of these five studies, all except [43], which measured oxygenated hemoglobin, employed EEG. Studies examining music [13] and running [42] supported the THH, whereas in the remaining three studies, none of the proposed theories received support (see Table 2 in the online appendix).

Absorption and Fluency. Two studies investigated second-order constructs in knowledge work [12] or gaming [52]. While [12] found quadratic relationships between beta power asymmetry and both absorption and fluency, [52] reported positive correlations between theta and delta power and absorption, but did not investigate fluency. Neither study provides a clear relation to any of the three theories.

Flow Dimensions (see Chapter 2, first paragraph)**.** One study examined all nine flow dimensions in music [13]. However, findings were only embedded in the THH at the level of global flow.

Cortisol measurement, eye tracking, and facial tracking

Cortisol. Only one study examined the relationship using a gaming task and found no association between cortisol and *global flow*. The authors conclude that there is no universal relation between cortisol and flow and therefore propose the "treatment specificity and personal characteristics hypothesis", suggesting that cortisol reactions depend on task and personal characteristics of the participants. In contrast, [23, 34, 53] investigated the relationship between cortisol and *absorption* and *fluency* in gaming tasks. Two studies indicate a (partially) inverse U-shaped relationship between cortisol levels and absorption [23, 34], but no association with fluency, whereas [53] identified a negative linear relationship for both fluency and absorption.

Eye Tracking. [44] examined the relationship between the quiet eye and *global flow* in athletes shooting free throws, indicating a significant negative linear association. [54] identified positive associations between the percentage of fixation duration and *time distortion* and negative associations between the percentage of fixation count and time distortion in a gaming context (see Table 3 in the online appendix).

Facial Tracking. Two studies investigated *global flow* in knowledge work tasks [45, 46], both indicating significant associations between flow and facial activity and subsequent derived emotions. Whereas [45] identified surprise and z-axis head orientation as significant negative correlates of flow among thesis-writing participants, though the sample of four considerably limits statistical power, [46] identified happiness expression as a significant negative correlate in a language learning class.

Discussion and Limitations

Our review integrates neurophysiological findings by focusing on studies grounded in Csikszentmihalyi's flow theory [1]. On this basis, we outline methodological and theoretical implications and delineate directions for future flow research.

Methodological implications

First, flow is highly task dependent, as even tasks within the same category, such as gaming, may differ in how they contribute to individual flow dimensions [11, 55]. For instance, the level of flow experienced by participants increased more gradually in Tetris than in an action game over the course of six playing sessions [55]. This heterogeneity limits comparability and the detection of consistent patterns across studies, as reflected in the wide range of neurophysiological findings. Addressing this issue requires a framework for categorizing tasks according to their flow-relevant characteristics. While [56] propose a promising approach based on four categories, namely *interaction speed*, *feedback*, *strategy*, and *goals*, to classify isolated characteristics of tasks; it does not integrate physical activity, which substantially influences physiological responses. As a result, the approach is only conditionally applicable to examine the relationship between flow and physiological activation.

- *Future research* should build on this groundwork to develop a classification framework for flow that enables cross-study comparison. Additionally, measuring physiological responses in the same individuals across multiple tasks could yield valuable insights into task-dependent differences.

Second, the analysis of flow dimensions has gained increasing importance, as not all dimensions contribute equally to the experience [5, 34]. Moreover, including antecedents, which are only the prerequisites, in the global flow calculation does not truly reflect the experience itself [56]. Findings of this review further support the need for dimension-level analysis. For absorption but not fluency, a significant positive relationship between HF-HRV and cortisol levels was identified [34], a pattern further supported by an inverse U-shaped relationship between cortisol levels and absorption, but not fluency, among chess players [23]. Similarly, an inverse U-shaped association with alpha neural oscillations was reported in the absorption but not for fluency [45]. One possible explanation is that absorption only emerges when perceived challenge and perceived skill are balanced, thus exclusively within the flow channel, and may therefore serve as a more reliable indicator for the flow experience [34].

- *Future research* should therefore include analysis of flow dimensions to advance the understanding of dimension-specific neurophysiological correlates of flow.

Third, the majority of the studies relied on a difficulty manipulation paradigm, varying between two [52], three [22, 28, 51], four [39], or five difficulty levels [53] to induce flow. Notably, [39] included an autonomy condition in which participants self-selected their difficulty level, with the instruction to select a level that is neither too easy nor too hard. Although the autonomy condition did not yield higher flow scores compared to the experimenter-defined flow condition, it resulted in greater physiological activations. This suggests that higher individual autonomy may produce more stimulant experiences, because flow is an intrinsic state influenced by the individual's motivation to perform the task [57].

While controlled laboratory environments are necessary to minimize the effect of confounding variables, particularly given the sensitivity of physiological measurement to external disturbances [58], researchers have called for field studies to complement laboratory findings [19, 40]. Only two studies included in the review were conducted as field experiments [12, 19], of which only one compared a natural task (own project work) with a mental arithmetic task in the field, reporting more intense flow experiences in the first [12]. In addition, none of the included studies employed a dual-task paradigm, which is applied to indirectly assess flow by measuring the level of attention. As complete absorption during flow hinders an individual from allocating attentional resources to irrelevant external stimuli, secondary task reaction times have been proposed as a reliable proxy measure for the flow state [11].

- *Future research* should consider integrating autonomy conditions to better reflect intrinsic motivation, conducting field studies with natural tasks to elicit *real* flow experiences, and further exploring the potential of dual-task paradigms.

Fourth, the definition of participant inclusion criteria presents another approach to ensure the elicitation of flow experiences. As experts are more prone to experience flow states in their domain of expertise [41], several studies examined whether expertise moderates the physiological correlates of flow. In a gaming context, less activation in frontal and parietal cortex regions was identified in non-gamers compared to gamers, attributed to a lower degree of automaticity, because their lack of skills requires a more explicit information processing system [49]. Similarly, the difference in upper alpha power between flow and non-flow states was more pronounced in musicians with higher expertise, suggesting that experts are more likely to rely on implicit, automatic processing than non-experts. However, findings only provided statistical significance after outlier removal [13]. Furthermore, significant differences in ventrolateral prefrontal cortex activation between gamers and non-gamers indicate that experts demonstrate better control mechanisms to regulate attention and emotion [21]. Collectively, these findings suggest that experience moderates neurophysiological activation during flow, and that the THH receives support for experts, but might differ for non-experts.

The flow experience is further affected by personality traits such as low neuroticism or high conscientiousness [59]. Despite calls to include screening for flow disposition [59, 60], only two studies in this review implement screening procedures [12, 13]. While no significant differences in flow proneness across participants were identified in the first study, limiting the scope for further analysis [12], the second reported that differences in theta band functional connectivity analysis between flow and non-flow were primarily reported in participants with high dispositional flow scores [13]. The authors link increased activity to superior attentional control in flow-prone individuals.

- *Future research* should systematically screen for expertise, flow proneness, and specify inclusion criteria for participants, as studies targeting experts and flow-prone individuals appear more likely to identify neurophysiological patterns related to flow.

Fifth, this review identifies promising directions for future non-obtrusive flow measurements, such as the use of eye and facial activity [44, 45]. Integrating these parameters into a camera-based approach, which requires no direct skin contact, could prevent the physical discomfort induced by wearables, such as fNIRS sensors [28]. In addition, cameras enable the remote assessment of heart rate (variability) [61], integrating well-established physiological flow parameters into a single non-intrusive device.

- *Future research* should therefore explore the potential of camera or webcam-based flow detection.

Theoretical implications

First, this review indicates that physiological activation changes across the temporal course of flow, including the time before, during, and after the flow state. [40] suggest that elevated SNS activity before entering the flow state may facilitate flow, whereas peak flow was observed during increased PSNS activation. This temporal dynamic is further emphasized by [11], who demonstrated that higher PSNS activity at the end of the task, compared to the beginning, might be associated with higher reported flow.

- *Future research* should examine the temporal dynamics of the flow state to contribute to a more precise understanding of the underlying relationship.

Second, the distinction between flow and stress has been intensively studied [8]. While earlier studies characterized the physiological activation as similar to stress responses with strainful tension and mental load [62], the findings of the present review indicate that these assumptions have since been refined. Restricting the analysis to studies employing a ground-truth measurement consistent with Csikszentmihalyi's flow theory, the PSNS activation emerges as a central component. The co-activation of both ANS branches has been highlighted across several studies [33, 34, 39, 41, 48, 49], supporting the assumption that flow is determined by maximum physiological flexibility, enabling adaptive responses to changing environmental demands [11], a pattern that distinguishes it from stress [22].

Third, the selection and interpretation of physiological parameters as indicators for ANS activation remain debated (see discussion on low-frequency heart rate variability in [63]), prompting calls for more reliable and valid measures [28, 49]. Addressing this gap, [19] integrated the pre-ejection period (PEP) in their analysis. PEP describes the time interval between the stimulation of the left ventricle and the opening of the aortic valve, and a shortening of this interval signals increased sympathetic activation. During live dancing performance, the change in PEP from baseline rest to performance explained 8.8% in variance in the *sense of control* dimension.

- *Future research* should therefore systematically screen for recent developments of physiological indicators for ANS activation, in particular drawing on advances from medical clinical research, and transfer findings to future flow research.

Fourth, across the included studies, four find support for the THH [13, 39, 42, 49]. However, [28] did not find a relation between frontal cortical oxygenation and flow, and thus specifically disagrees with flow as a state of hypofrontality. While [28] employed fNIRS measurements, which offer higher spatial but lower temporal resolution, the remaining studies relied on EEG, which offers higher temporal but lower spatial resolution [64], suggesting that the mixed findings regarding THH support may reflect methodological differences between studies. Notably, several studies lack an embedding of empirical findings within explicit theoretical mechanisms (e.g., [43, 52]), constraining current theory development.

- *Future research* should continue to employ high spatial and temporal resolution neurophysiological methods and embed findings within theoretical frameworks to advance theory development.

Fifth, significant relationships could be identified across all five frequency bands based on EEG activity. However, only [12] identified quadratic relationships with beta and theta activity, while the remaining studies focused on linear relations. For example, two studies [15, 52] report positive linear relations in central and parietal regions, interpreting flow as a state of high concentration and cognitive engagement. Furthermore, a state of relaxation is derived from this relationship [52]. Delta and theta bands were further identified as the most robust features for flow prediction, outperforming predictive modelling for the combined five frequency bands [15]. In contrast, theta activity did not significantly contribute in a multivariate model, suggesting that other frequency bands may play a more prominent role in characterizing the flow state [14].

- *Future research* should systematically examine the relative contributions of individual frequency bands, using approaches to disentangle their independent and combined roles in characterizing the flow state.

Limitations

Within this review, a common ground truth was applied to establish a basis for comparison across studies. While we refer to scales that rely on Csikszentmihalyi's flow theory as the ground truth, this does not imply their validity. An example is the Flow State Scale 2, which does not sufficiently discriminate between flow and clutch, a state in which an individual is successful in pressured situations, which partially overlaps with the flow experience [65]. Additionally, global flow scores calculated by averaging antecedent, characteristics, and consequence dimensions may inadequately represent the flow state [3, 56].

- *Future research* should therefore prioritize the evaluation of flow scales, with particular attention to their discriminant validity.

Concluding Remarks

This review reveals heterogeneity and conceptual fragmentation in the neurophysiological correlates of flow. Nevertheless, the findings demonstrate the potential of neurophysiological measures for flow detection and for informing future flow-supportive systems. Specifically, we outline methodological and theoretical implications, delineate directions for future research, and call for stronger integration of neurophysiological findings into existing theoretical accounts to advance theory development in the field.

References

1. Csikszentmihalyi, M.: Beyond boredom and anxiety. Jossey-Bass Publ, San Francisco, Calif. (1975)
2. Rissler, R., Nadj, M., Adam, M.: Flow in Information Systems Research: Review, Integrative Theoretical Framework, and Future Directions. In: Leimeister, J.M., Brenner, W. (eds.) Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), pp. 1051–1065. St. Gallen
3. Moneta, G.B.: On the Conceptualization and Measurement of Flow. In: Peifer, C., Engeser, S. (eds.) Advances in Flow Research, pp. 31–69. Springer International Publishing, Cham (2021). doi: 10.1007/978-3-030-53468-4_2
4. Peifer, C., Syrek, C., Ostwald, V., Schuh, E., Antoni, C.H.: Thieves of Flow: How Unfinished Tasks at Work are Related to Flow Experience and Wellbeing. J Happiness Stud, vol. 21, 1641–1660 (2020). doi: 10.1007/s10902-019-00149-z
5. Nadj, M., Rissler, R., Adam, M., Knierim, M.T., Li, M., Maedche, A., Riedl, R.: What Disrupts Flow in Office Work? The Impact of Frequency and Relevance of IT-Mediated Interruptions. MISQ, vol. 47, 1615–1646 (2023). doi: 10.25300/MISQ/2023/17200
6. Rissler, R., Nadj, M., Li, M.X., Loewe, N., Knierim, M.T., Maedche, A.: To Be or Not to Be in Flow at Work: Physiological Classification of Flow Using Machine Learning. IEEE Trans. Affective Comput., vol. 14, 463–474 (2020). doi: 10.1109/TAFFC.2020.3045269
7. Knierim, M.T., Pieper, V., Schemmer, M., Loewe, N., Reali, P.: Predicting In-Field Flow Experiences Over Two Weeks from ECG Data: A Case Study. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 52, pp. 96–102. Springer International Publishing, Cham (2021). doi: 10.1007/978-3-030-88900-5_11
8. Knierim, M.T., Rissler, R., Dorner, V., Maedche, A., Weinhardt, C.: The Psychophysiology of Flow: A Systematic Review of Peripheral Nervous System Features. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 25, pp. 109–120. Springer International Publishing, Cham (2017). doi: 10.1007/978-3-319-67431-5_13
9. Alameda, C., Sanabria, D., Ciria, L.F.: The brain in flow: A systematic review on the neural basis of the flow state. Cortex, vol. 154, 348–364 (2022). doi: 10.1016/j.cortex.2022.06.005
10. Nah, F.F.-H., Yelamanchili, T., Siau, K.: A Review on Neuropsychophysiological Correlates of Flow. In: Nah, F.F.-H., Tan, C.-H. (eds.) HCI in Business, Government and Organizations. Supporting Business. Lecture Notes in Computer Science, vol. 10294, pp. 364–372. Springer International Publishing, Cham (2017). doi: 10.1007/978-3-319-58484-3_28
11. Khoshnoud, S., Alvarez Igarzábal, F., Wittmann, M.: Peripheral-physiological and neural correlates of the flow experience while playing video games: a comprehensive review. PeerJ, vol. 8, e10520 (2020)

12. Knierim, M.T., Stano, F., Kurz, F., Heusch, A., Wilson, M.L.: Exploring Flow in Real-World Knowledge Work Using Discrete cEEGrid Sensors. In: Yamashita, N., Evers, V., Yatani, K., Ding, X., Lee, B., Chetty, M., Toups-Dugas, P. (eds.) Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, pp. 1–16. ACM, New York, NY, USA (2025). doi: 10.1145/3706598.3713512
13. Tan, J., Di Bernardi Luft, C., Bhattacharya, J.: The After-Glow of Flow: Neural Correlates of Flow in Musicians. Creativity Research Journal, vol. 36, 469–490 (2024). doi: 10.1080/10400419.2023.2277042
14. Lin, B., Guo, B., Zhuang, L., Zhang, D., Wang, F.: Neural oscillations predict flow experience. Cognitive neurodynamics, vol. 19, 1 (2024). doi: 10.1007/s11571-024-10205-x
15. Hang, Y., Unenbat, B., Tang, S., Wang, F., Lin, B., Zhang, D.: Exploring the Neural Correlates of Flow Experience with Multifaceted Tasks and a Single-Channel Prefrontal EEG Recording. Sensors (Basel, Switzerland), vol. 24 (2024). doi: 10.3390/s24061894
16. Khanganba, S.P., Ubaradka, A.: Immersion in the Shadows: Physiological Markers of Flow Experience in a Virtual Reality Survival Horror Game. In: Jin, S., Kim, J.H., Kong, Y.-K., Park, J., Yun, M.H. (eds.) Proceedings of the 22nd Congress of the International Ergonomics Association, Volume 4. Springer Series in Design and Innovation, vol. 56, pp. 454–460. Springer Nature Singapore, Singapore (2025). doi: 10.1007/978-981-95-0289-9_67
17. Ulrich, M., Keller, J., Grön, G.: Dorsal Raphe Nucleus Down-Regulates Medial Prefrontal Cortex during Experience of Flow. Frontiers in behavioral neuroscience, vol. 10, 169 (2016). doi: 10.3389/fnbeh.2016.00169
18. Agarwal, R., Karahanna, E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage1. MISQ, vol. 24, 665–694 (2000). doi: 10.2307/3250951
19. Jaque, S.V., Thomson, P., Zaragoza, J., Werner, F., Podeszwa, J., Jacobs, K.: Creative Flow and Physiologic States in Dancers During Performance. Frontiers in psychology, vol. 11, 1000 (2020). doi: 10.3389/fpsyg.2020.01000
20. Nakamura, J., Csikszentmihalyi, M.: Flow Theory and Research. In: Lopez, S.J., Snyder, C.R., Nakamura, J., Csikszentmihalyi, M. (eds.) The Oxford Handbook of Positive Psychology, pp. 194–206. Oxford University Press (2009). doi: 10.1093/oxfordhb/9780195187243.013.0018
21. Yu, D., Wang, S., Song, F., Liu, Y., Zhang, S., Wang, Y., Xie, X., Zhang, Z.: Research on user experience of the video game difficulty based on flow theory and fNIRS. Behaviour & Information Technology, vol. 42, 789–805 (2022). doi: 1080/0144929X.2022.2043442
22. Tozman, T., Magdas, E.S., MacDougall, H.G., Vollmeyer, R.: Understanding the psychophysiology of flow: A driving simulator experiment to investigate the relationship between flow and heart rate variability. Computers in Human Behavior, vol. 52, 408–418 (2015). doi: 10.1016/j.chb.2015.06.023
23. Tozman, T., Zhang, Y.Y., Vollmeyer, R.: Inverted U-Shaped Function Between Flow and Cortisol Release During Chess Play. J Happiness Stud, vol. 18, 247–268 (2017). doi: 10.1007/s10902-016-9726-0
24. Riedl, R., Léger, P.-M.: Tools in NeuroIS Research: An Overview. In: Riedl, R., Léger, P.-M. (eds.) Fundamentals of NeuroIS. Studies in Neuroscience, Psychology and

Behavioral Economics, pp. 47–72. Springer Berlin Heidelberg, Berlin, Heidelberg (2016). doi: 10.1007/978-3-662-45091-8_3
25. Waxenbaum, J.A., Vamsi Reddy, Joe M. Das: Anatomy, Autonomic Nervous System. In: Waxenbaum, J.A., Reddy, V., Das, J.M. (eds.) StatPearls [Internet]. StatPearls Publishing (2025)
26. Berntson, G.G., Norman, G.J., Hawkley, L.C., Cacioppo, J.T.: Cardiac autonomic balance versus cardiac regulatory capacity. Psychophysiology, vol. 45, 643–652 (2008). doi: 10.1111/j.1469-8986.2008.00652.x
27. Keller, J.: The Flow Experience Revisited: The Influence of Skills-Demands-Compatibility on Experiential and Physiological Indicators. In: Harmat, L., Ørsted Andersen, F., Ullén, F., Wright, J., Sadlo, G. (eds.) Flow Experience, pp. 351–374. Springer International Publishing, Cham (2016). doi: 10.1007/978-3-319-28634-1_21
28. Harmat, L., Manzano, Ö. de, Theorell, T., Högman, L., Fischer, H., Ullén, F.: Physiological correlates of the flow experience during computer game playing. International journal of psychophysiology, vol. 97, 1–7 (2015). doi: 10.1016/j.ijpsycho.2015.05.001
29. Dietrich, A.: Neurocognitive mechanisms underlying the experience of flow. Consciousness and cognition, vol. 13, 746–761 (2004). doi: 10.1016/j.concog.2004.07.002
30. Weber, R., Tamborini, R., Westcott-Baker, A., Kantor, B.: Theorizing Flow and Media Enjoyment as Cognitive Synchronization of Attentional and Reward Networks. Communication Theory, vol. 19, 397–422 (2009). doi: 10.1111/j.1468-2885.2009.01352.x
31. Gold, J., Ciorciari, J.: A neurocognitive model of flow states and the role of cerebellar internal models. Behavioural brain research, vol. 407, 113244 (2021). doi: 10.1016/j.bbr.2021.113244
32. Hiller-Sturmhöfel, S., Bartke, A.: The endocrine system: an overview. Alcohol health and research world, vol. 22, 153–164 (1998)
33. Bian, Y., Yang, C., Gao, F., Li, H., Zhou, S., Li, H., Sun, X., Meng, X.: A framework for physiological indicators of flow in VR games: construction and preliminary evaluation. Pers Ubiquit Comput, vol. 20, 821–832 (2016). doi: 10.1007/s00779-016-0953-5
34. Peifer, C., Schulz, A., Schächinger, H., Baumann, N., Antoni, C.H.: The relation of flow-experience and physiological arousal under stress — Can u shape it? Journal of Experimental Social Psychology, vol. 53, 62–69 (2014). doi: 10.1016/j.jesp.2014.01.009
35. Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., Glenk, L.M.: Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. Computers & Education, vol. 79, 69–100 (2014). doi: 10.1016/j.compedu.2014.07.001
36. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, vol. , xiii–xxiii (2002)
37. Swann, C., Keegan, R.J., Piggott, D., Crust, L.: A systematic review of the experience, occurrence, and controllability of flow states in elite sport. Psychology of Sport and Exercise, vol. 13, 807–819 (2012). doi: 10.1016/j.psychsport.2012.05.006
38. Tan, L., Sin, H.X.: Flow research in music contexts: A systematic literature review. Musicae Scientiae, vol. 25, 399–428 (2021). doi: 10.1177/1029864919877564
39. Sampaio Barros, M.F. de, Araújo-Moreira, F.M., Trevelin, L.C., Radel, R.: Flow experience and the mobilization of attentional resources. Cognitive, affective & behavioral neuroscience, vol. 18, 810–823 (2018). doi: 10.3758/s13415-018-0606-4

40. Jha, S., Stogios, N., Oliveira, A.S. de, Thomas, S., Nolan, R.P.: Getting Into the Zone: A Pilot Study of Autonomic-Cardiac Modulation and Flow State During Piano Performance. Frontiers in psychiatry, vol. 13, 853733 (2022). doi: 10.3389/fpsyt.2022.853733
41. De Manzano, O., Theorell, T., Harmat, L., Ullén, F.: The psychophysiology of flow during piano playing. Emotion (Washington, D.C.), vol. 10, 301–311 (2010). doi: 10.1037/a0018432
42. Wollseiffen, P., Schneider, S., Martin, L.A., Kerhervé, H.A., Klein, T., Solomon, C.: The effect of 6 h of running on brain activity, mood, and cognitive performance. Experimental brain research, vol. 234, 1829–1836 (2016). doi: 10.1007/s00221-016-4587-7
43. Takinami, S., Nakatani, Y., Ono, Y., Murakami, S.: Contribution of flow state to the discrepancy between perceived exercise intensity and physiological intensity during endurance bike exercise. BMC sports science, medicine & rehabilitation, vol. 17, 273 (2025). doi: 10.1186/s13102-025-01342-9
44. Harris, D.J., Vine, S.J., Wilson, M.R.: Flow and quiet eye: the role of attentional control in flow experience. Cognitive processing, vol. 18, 343–347 (2017). doi: 10.1007/s10339-017-0794-9
45. Berger, C., Knierim, M.T., Weinhardt, C.: Detecting Flow Experiences in the Field Using Video-Based Head and Face Activity Recognition: A Pilot Study. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.-M., Randolph, A.B., Müller-Putz, G. (eds.) Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation, vol. 52, pp. 120–127. Springer International Publishing, Cham (2021). doi: 10.1007/978-3-030-88900-5_14
46. Hu, X., Gao, J.: Facial expression recognition reveals students' engagement in online class: Correlations with six engagement measurements. PloS one, vol. 20, e0334232 (2025). doi: 10.1371/journal.pone.0334232
47. Mansfield, B.E., Oddson, B.E., Turcotte, J., Couture, R.T.: A possible physiological correlate for mental flow. The Journal of Positive Psychology, vol. 7, 327–333 (2012). doi: 10.1080/17439760.2012.691982
48. Chin, M.S., Kales, S.N.: Is There an Optimal Autonomic State for Enhanced Flow and Executive Task Performance? Frontiers in psychology, vol. 10, 1716 (2019). doi: 10.3389/fpsyg.2019.01716
49. Khoshnoud, S., Alvarez Igarzábal, F., Wittmann, M.: Brain-Heart Interaction and the Experience of Flow While Playing a Video Game. Frontiers in human neuroscience, vol. 16, 819834 (2022). doi: 10.3389/fnhum.2022.819834
50. Wolf, S., Brölz, E., Keune, P.M., Wesa, B., Hautzinger, M., Birbaumer, N., Strehl, U.: Motor skill failure or flow-experience? Functional brain asymmetry and brain connectivity in elite and amateur table tennis players. Biological psychology, vol. 105, 95–105 (2015). doi: 10.1016/j.biopsycho.2015.01.007
51. Su, K.-C., Wu, K.-C., Chou, K.-R., Huang, C.-H.: Tongue Muscle Training App for Middle-Aged and Older Adults Incorporating Flow-Based Gameplay: Design and Feasibility Pilot Study. JMIR serious games, vol. 13, e53045 (2025). doi: 10.2196/53045
52. Metin, B., Göktepe, A.K., Sutcubasi, B., Serin, E., Tas, C., Dolu, F., Tarhan, K.N.: EEG findings during flow state. JNBS, vol. , 1 (2017). doi: 10.5455/JNBS.1496152464

53. Peifer, C., Schächinger, H., Engeser, S., Antoni, C.H.: Cortisol effects on flow-experience. Psychopharmacology, vol. 232, 1165–1173 (2015). doi: 10.1007/s00213-014-3753-5
54. Tsai, M.-J., Huang, L.-J., Hou, H.-T., Hsu, C.-Y., Chiou, G.-L.: Visual behavior, flow and achievement in game-based learning. Computers & Education, vol. 98, 115–129 (2016). doi: 10.1016/j.compedu.2016.03.011
55. Belchior, P., Marsiske, M., Sisco, S., Yam, A., Mann, W.: Older adults' engagement with a video game training program. Activities, adaptation & aging, vol. 36, 269–279 (2012). doi: 10.1080/01924788.2012.702307
56. Durcan, O., Holland, P., Bhattacharya, J.: A framework for neurophysiological experiments on flow states. Communications psychology, vol. 2, 66 (2024). doi: 10.1038/s44271-024-00115-3
57. Moller, A.C., Meier, B.P., Wall, R.D.: Developing an Experimental Induction of Flow : Effortless Action in the Lab. In: Bruya, B. (ed.) Effortless Attention, pp. 191–204. The MIT Press (2010). doi: 10.7551/mitpress/9780262013840.003.0010
58. Fairclough, S.H.: Fundamentals of physiological computing. Interacting with Computers, vol. 21, 133–145 (2009). doi: 10.1016/j.intcom.2008.10.011
59. Ullén, F., Manzano, Ö. de, Almeida, R., Magnusson, P.K., Pedersen, N.L., Nakamura, J., Csíkszentmihályi, M., Madison, G.: Proneness for psychological flow in everyday life: Associations with personality and intelligence. Personality and Individual Differences, vol. 52, 167–172 (2012). doi: 10.1016/j.paid.2011.10.003
60. Wonders, M., Hodgson, D., Whitton, N.: Measuring Flow: Refining Research Protocols That Integrate Physiological and Psychological Approaches. Human Behavior and Emerging Technologies, vol. 2025 (2025). doi: 10.1155/hbe2/6464984
61. Xiao, H., Liu, T., Sun, Y., Li, Y., Zhao, S., Avolio, A.: Remote photoplethysmography for heart rate measurement: A review. Biomedical Signal Processing and Control, vol. 88, 105608 (2024). doi: 10.1016/j.bspc.2023.105608
62. Keller, J., Bless, H., Blomann, F., Kleinböhl, D.: Physiological aspects of flow experiences: Skills-demand-compatibility effects on heart rate variability and salivary cortisol. Journal of Experimental Social Psychology, vol. 47, 849–852 (2011). doi: 10.1016/j.jesp.2011.02.004
63. Berntson, G.G., Bigger, J.T., Eckberg, D.L., Grossman, P., Kaufmann, P.G., Malik, M., Nagaraja, H.N., Porges, S.W., Saul, J.P., Stone, P.H., et al.: Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology, vol. 34, 623–648 (1997). doi: 10.1111/j.1469-8986.1997.tb02140.x
64. Li, R., Yang, D., Fang, F., Hong, K.-S., Reiss, A.L., Zhang, Y.: Concurrent fNIRS and EEG for Brain Function Investigation: A Systematic, Methodology-Focused Review. Sensors (Basel, Switzerland), vol. 22 (2022). doi: 10.3390/s22155865
65. Jackman, P.C., Crust, L., Swann, C.: Systematically comparing methods used to study flow in sport: A longitudinal multiple-case study. Psychology of Sport and Exercise, vol. 32, 113–123 (2017). doi: 10.1016/j.psychsport.2017.06.009

AI-Assisted Learning of Integrated Business Processes in SAP: A Proposed Eye-Tracking Study of Cognitive Engagement Strategies

Palash Bera,[1] Sandip Sarkar[1] Deepika Gopukumar[1],

[1] Chaifetz School of Business, Saint Louis University, Saint Louis, USA
Palash.bera@slu.edu, sandipkumar.sarkar@slu.edu,
deepika.gopukumar.1@slu.edu

Abstract. This proposed research study investigates whether AI assistance can help learning of integrated business processes from SAP screens and examines the cognitive mechanisms underlying such learning. While prior studies have demonstrated AI's potential to enhance learning outcomes, the cognitive processes through which learners engage with both AI and visual information remain underexplored. By combining experimental manipulation of AI assistance with eye-tracking methodology, this study aims to uncover how learners allocate visual attention, formulate questions, and develop mental models of business process integration through SAP screens.

Keywords: AI Assisted learning, GenAI, Eye tracking, SAP, Cognitive behavior.

Introduction

The rise of generative AI (GenAI) is transforming how users in complex domains learn and acquire knowledge. ERP systems integrate different business units and automate business activities for decision-making [1]. However, learning ERP requires a different set of skills compared to traditional technologies [2]. ERP requires business domain knowledge in addition to the technical knowledge, as users need to understand how multiple organizational units impact business operations [3].

In ERP systems such as SAP, business processes involve different business functions such as procurement, sales, and accounting, with transactions in one screen creating visible effects across multiple interrelated screens. For novice SAP users, understanding the dependencies of such business processes is critical for making sound business decisions. Traditional learning approaches often struggle to convey this integration effectively, as learners must mentally connect information distributed across multiple screens while grasping the underlying business concepts.

This research investigates how AI can help learning of integrated business processes from related SAP screens and critically examines the cognitive mechanisms underlying such learning. While prior studies have demonstrated AI's potential to enhance learning outcomes, the cognitive processes through which learners engage with both AI and visual information remain underexplored [4]. By combining AI assistance with eye-tracking methodology, this study aims to uncover how learners allocate visual attention, formulate questions, and develop cognitive strategies to solve SAP related problems.

This study is positioned within the field of NeuroIS which is the application of neurophysiological tools and theories to information systems research [5-7]. While prior NeuroIS research has examined phenomena such as trust, cognitive load, and technology adoption, less attention has been given to the *cognitive mechanisms or strategies of learning* with GenAI tools. By using eye-tracking as a physiological measure of attention and integration, we contribute to NeuroIS by revealing how the quality of human AI interaction shapes learning outcomes. Following Riedl & Léger [8], we consider eye movements as an objective window into otherwise unobservable cognitive processes (e.g., selection, organization, and integration).

Related work

Learning and Generative AI

Generative AI has transformed personalized learning support by enabling natural dialogue where learners can ask questions about specific concepts and cross-screen relationships. Baek and Kim [4] identified five key motivations for ChatGPT usage—information seeking, task efficiency, personalization, social interaction, and playfulness, finding that personalization and task efficiency positively impact trust and continuance intention. Marimon et al. [9] demonstrated that GenAI adoption improves performance only through trust as a mediating variable, highlighting that psychological factors may be more critical than just access to GenAI. Zhang et al. [10] distinguish between cognitive use (knowledge acquisition) and social use (collaboration) of GenAI, finding both usages enhance performance through knowledge transfer and resource acquisition. However, Zhang [11] warns the risk of cognitive overreliance of GenAI as users increasingly delegate cognitive tasks such as summarization, coding, or ideation to GPT models. Such usage may lead to a gradual erosion of independent reasoning skills.

Learning Integrated Business Processes using SAP

SAP is one of the most recognized and popular ERP systems used worldwide. Learning SAP systems presents unique challenges because business processes are distributed across multiple interconnected screens, requiring learners to recognize recurring identifiers and to understand how transactions in multiple business functions such as procurement and sales affect one another [12]. While pedagogical approaches such as hands-on exercises and ERPsim simulations are commonly used in teaching business process integration, these methods have not been systematically studied using cognitive methodology such as eye-tracking.

Research question and model

Learning the use of SAP is complex as the business processes are shown in multiple different interrelated SAP screens. Problems such as fixing errors in a specific SAP screen require a deep understanding of how a screen is related to other screens. Moreover, conceptual knowledge of business processes is required to understand the relationship among the screens.

To help novice users solve complex SAP problems (e.g. fix errors in SAP screens), we propose an AI assisted tool that does not provide the answers to such problems but help the learners to understand SAP concepts and their relationships. Such knowledge in turn can help solve such problems. There are many AI-assisted tools such as automated writing evaluation systems, conversational chatbots, and adaptive learning platforms available that can provide personalized and flexible opportunities to learn [13]. Our focus is only on the conversational AI chatbot that is developed to support leaners providing in-depth SAP and business knowledge about a specific domain.

We propose that the use of a conversational AI chatbot (we refer to this as GenAI tool) influences learning outcomes through its effect on cognitive processing. The broad research question we investigate is *how does the quality of human-AI interaction affect learning outcomes?*

Fig. 16. Research Model

As illustrated in Figure 1, the independent variable is the use of a customized GenAI tool while performing SAP tasks which is hypothesized to affect learning outcomes via use of cognitive engagement strategies. By learning outcomes, we refer to the quality of the answers to problem solving questions on SAP that tests deep level understanding of SAP concepts. In the next section, we elaborate on the cognitive engagement strategies.

Cognitive Engagement Strategies

We propose that high quality learning depends on the quality of cognitive engagement and not simply on the availability of GenAI tool. Users need to employ specific cognitive strategies to engage with the GenAI tool to have high quality learning. We name this as *cognitive engagement strategies* and define it *as the deliberate, goal-directed cognitive actions a learner takes to process, transform, and integrate new information with existing knowledge*. This definition is based on the theories mentioned in the next sub section.

Related Theories

Cognitive engagement strategies are based on three complementary theories: Cognitive Load Theory [14], Cognitive Theory of Multimedia Learning (CTML) [15], and Generative Learning Theory [16].

Cognitive Load Theory [14] states that learning is constrained by the limited capacity of working memory. Two types of cognitive loads are relevant for this research. First, the intrinsic load which is the inherent difficulty of the material itself and is determined by the number of interacting information elements that must be held in working memory simultaneously [14]. Understanding an integrated business process in SAP (such as tracing a quantity discrepancy from a purchase order to a goods receipt to an invoice receipt) involves high intrinsic load because multiple screens and their numerical relationships must be processed together. Second, the extraneous load that is created by the instructions. In our study, extraneous load can arise as the SAP related problem-solving questions are quite complex and require to be broken into small manageable parts. Extraneous load can be reduced by altering instructional procedures, formats, or guidance [17]. As this type of load does not contribute to learning therefore it should be minimized.

CTML [15] proposes that meaningful learning occurs through three cognitive processes- selecting relevant information from the presented material, organizing it into coherent mental representations, and integrating those representations with prior knowledge retrieved from long-term memory. These processes are constrained by the limited capacity of each channel. CTML also predicts that learners who spontaneously perform these three processes will achieve better problem-solving outcomes.

Generative Learning Theory [16, 18] complements CLT and CTML by focusing on the learner's active construction of meaning. Generative learning involves "making sense" of provided learning material by actively organizing and integrating it with one's existing knowledge [18]. The intended outcome of generative learning is to construct coherent mental representations that enable learners to apply their knowledge to new situations [16]. Meaningful learning depends on the learner engaging in appropriate cognitive processing involving selecting, organizing, and integrating [15]. Generative learning activities prompt learners to make sense of material such as summarizing, self-explaining, drawing, and asking questions [19]. A question is a generative act because it requires the learner to detect a gap in their mental model, formulate a precise query, and actively seek information to fill that gap [19]. In the next sub-section, we suggest how cognitive engagement strategies can be implemented using these theories.

Using Cognitive Engagement Strategies

We propose that cognitive engagement strategies are of two types: *low-quality* engagement strategy and *high-quality* engagement strategy. By providing access to the GenAI tool does not automatically guarantee high quality learning as it is up to the learners to implement cognitive engagement strategies. Fiorella [16] however mentions that unfortunately many learners do not spontaneously engage in sense-making behavior resulting in shallow learning. In the low-quality engagement strategy, learners ask factual questions that retrieve directly information that is seen on the screen. Such questions reflect passive processing but do not require the learner to construct new relationships or resolve inconsistencies. This strategy may lead to shallow learning.

In the high-quality engagement strategy, effective learning occurs from multiple sources of complex information (e.g., several SAP screens). Implementation of this engagement strategy will lower cognitive loads (both intrinsic and extraneous) and can be achieved by engaging in generative learning activity in the form of asking questions to make sense of the material. Learners can employ selection strategies that focus on key information fields that are critical for understanding the underlying problem-solving question and then use integration strategy to mentally (or visually) relate those fields across different screens. Learners who deliberately compare and connect information across screens are engaging in deeper integration, which is necessary to grasp cross-screen dependencies.

Both strategies can be implemented using the questions mentioned in Table 1. Examples of such questions related to SAP are provided on the table. Note that factual questions that focus on understanding the key concept to understand the underlying business process can be considered as part of high engagement strategy if such questions are followed up by cross-sectional questions. Learners who do not have access to such GenAI tool can still have high quality learning outcome, provided they are able to engage with SAP screens and employ high quality cognitive engagement strategies.

Table 10. Cognitive engagement strategies implementation

Question type	Strategy implemented	Example
Cross-screen relational - deliberately comparing and connecting information across screens and engaging in deep integration	High quality	How does the 'quantity received' on the goods receipt screen relate to the 'invoice quantity' on the invoice receipt screen?
Follow-up conditional- formulating hypotheses and seeking the evidence of support of them	High quality	If the goods receipt shows 450 units but the invoice says 500, what happens to the financial statement accounts?
Factual (screen-based)- seeking understanding of the meaning of concepts and terms	Low quality	What is the unit price of milk on the purchase order screen?

Hypotheses

We propose a set of hypotheses to test the effectiveness of cognitive engagement strategies using a GenAI tool. Learners who ask high-quality generative questions particularly cross-screen relational and follow-up conditional questions will experience reduced extraneous and intrinsic loads and implement cognitive processes as mentioned in CTML leading to superior learning outcomes.

H1: Learners who ask more specific screen-based questions, cross-screen questions, and follow up questions will perform better in SAP related problem-solving questions.

H2: Question-asking and screen viewing patterns will differ between high-performing and low-performing learners in the GenAI tool group, with high performers asking more cross-screen relational and follow-up conditional questions and visualizing the corresponding part of the screens where such cross sectional and follow up concepts are mentioned.

The use of high-quality cognitive strategies can also be employed by learners without the use of GenAI tool. For these learners, effective attention-based strategies are self-directed, and they must independently decide which SAP fields/screens matter and how they relate to each other. Those learners who do not look at the right parts of the relevant screens would perform poorly showing low cognitive engagement strategy.

H3: Learners without the availability of GenAI tool will perform better in answering the problem-solving questions if they view key fields of screens that establish cross-screen connections related to the SAP related problem-solving questions.

Methodology and Procedure

Eye Tracking

Eye-tracking is a well-validated neurophysiological tool for measuring attention, cognitive load, and information integration [20]. Eye-tracking has emerged as a powerful tool for studying cognitive processes in psychology and Information Systems studies [21]. Eye metrics provide objective insights into attentional focus. For example, prolonged fixation durations measured as the time spent gazing at task-critical elements are indicative of deliberate mindful processing, while reduced saccadic frequency (fewer rapid eye movements between unrelated areas) reflects lower distraction levels [22]. While eye-tracking has been applied to ERP usability [23], no work has integrated ERP, GenAI, and eye-tracking to examine how use of GenAI drive attentional behaviors influencing ERP task performance.

The use of eye-tracking in this study follows established NeuroIS practices [6]. In our study, eye-tracking operationalizes the constructs of selection (fixations on key fields), organization (scanpaths within a screen), and integration (gaze transitions across screens). This aligns with the NeuroIS research agenda that calls for linking brain/body measurements to established IS constructs [7].

Procedure

The proposed study will use a controlled laboratory experiment using eye-tracking to investigate how GenAI assistance influences cognitive processing and learning outcomes in SAP tasks. A between-subjects design will be used with two groups (30 in each)- an experimental group with access to a GenAI tool and a control group without the tool. Participants will be undergraduate business students enrolled in an introductory technology management course where SAP concepts are taught.

Participants will first fill up a survey on their existing domain (business concepts related to procurement, sales, and accounting) and SAP knowledge and then complete a set of comprehension questions (e.g. "The goods receipt and invoice receipt screen both reference 500 units of milk from V04") based on the SAP screens involving procurement, sales, and accounting transactions. The purpose of this exercise is to make the subjects familiar with all the SAP screens. This will be followed by problem-solving questions to measure learning outcomes. Subjects eye movements will be recorded while they perform the tasks. The answers to the problem-solving questions are not provided directly in the screens and subjects need to understand the interrelations of multiple screens to answer them. After answering these questions, subjects will answer questions related to trust and usage of GenAI.

Experimental Materials

Two sets of websites are designed to test the hypotheses. Both websites will have a set of static SAP screens related to the operations of a Milk distribution firm related to procurement, sales, and accounting. The effects of certain transactions (e.g. purchase of milk from a supplier) are reflected in these screens. Some screens such as financial statements show the status before and after the transaction has taken place. One website (Figure 2) will have an embedded GenAI chat tool that can answer questions related to any concepts mentioned in the screens or general business domain and SAP knowledge. The chat box is connected to GPT 4.0 from OpenAI where a knowledge base is created on the information on SAP screens and generic knowledge of SAP. A system prompt is customized to provide the answers related to the knowledge base. To answer the questions prompted by users, the knowledge base will be used first (to answer specific questions on SAP screens) and then supplemented with general knowledge of GPT 4.0. The prompt disallows subjects to provide answers to the problem-solving questions or modification to the questions.

We plan to use three problem solving questions such as "c*ompany 140 has created a purchase order for 500 units of milk with a freight charge of €1,000 as shown on the screen. However, assume that the supplier increases the freight charge to €2,000. How would this increase in freight charge affect the screens?"* The AI chat will not provide answer to this question if asked as is or if the numbers are modified (e.g. freight charge of €100 instead of €1,000). However, the AI chat allows asking questions such as what is the effect of high freight charges on business financials?

The chat interactions (conversations) are not used for training the GPT 4.0. This prevents the GPT from learning and providing answers to participants, as multiple participants can ask similar questions. All the interactions are captured in sessions that can be extracted after a subject completes the study. The second website has the same information as the first one, except the GenAI tool is not present. By using the exact website without the GenAI tool, we can measure the effect of the presence of the GenAI tool.

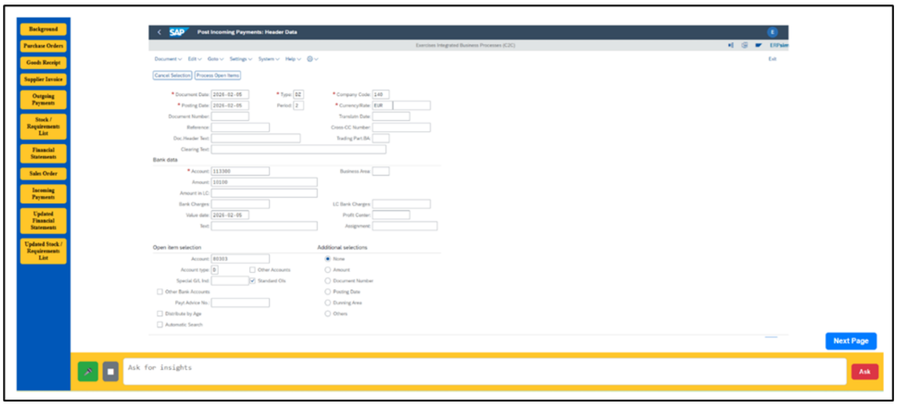


Fig. 2. Screenshot of the website with GenAI tool

Analysis Plan

Independent sample t-tests will compare problem-solving scores between groups. Domain and SAP experts (unaware of research objectives) will evaluate the quality of the problem-solving answers and provide a range of numerical scores to the problem-solving answers. SAP and domain prior knowledge and variables such as trust and usage of GenAI will be used as control variables. The interactions with the GenAI tool will be analyzed and the questions asked will be categorized as per Table 1. Analysis of the problem-solving answers and the types of questions asked will be used to test hypothesis one.

A key innovation in this research is the integration of eye-tracking and interaction data to examine cognitive processes. Several eye movement metrics such as fixation duration and fixation count will be used to indicate the depth of information processing and attention allocation to specific screen elements. The gaze transitions between areas of interest (AOIs) can reveal how learners integrate information across interconnected SAP screens. Key AOIs will include quantities, amounts, and supplier/customer identifiers that establish cross-screen connections in problem solving questions. The eye movements for answering cross sectional, factual, and follow up questions will be captured. For example, when a subject asks about a specific screen field, we can examine whether he/she was fixating on that field immediately before asking. We will also identify whether the subjects went back to the screen to verify the GenAI provided answers. The analysis of problem-solving performance, screen viewing patterns, and interactions will be used to test hypothesis two. To test the third hypothesis, problem-solving performance and the screen viewing patterns will be used.

Possible Contributions and conclusion

This research can make several novel contributions in the field of understanding the cognitive processes of using GenAI tools. First, it focuses on understanding the mechanism of implementing cognitive strategies using a customized GenAI tool. Second, it focuses on identifying when the use of these strategies leads to high performance. For this purpose, it uses combination of several data sources including eye tracking. Third, it compares two groups of users (with and without the use of GenAI) and identifies when high performance can be achieved even without the use of GenAI. Practically, the findings can inform organizations on how AI can be used to reduce cognitive load of novice SAP users.

This research can also highlight the role of asking questions related to cognitive engagement strategies. Such questions can be converted to theory-based prompt interventions embedded within the AI chat interface. Future research can test the effectiveness of these prompt interventions in providing high quality learning.

References

[1] M. I. Wijaya, "A decade of ERP teaching practice: A systematic literature review," *Education and Information Technologies,* pp. 1-21, 2023.

[2] J.-S. Park, Y.-S. Cho, and J. Koh, "A study on users' resistance toward ERP in the pre-adoption context," *Asia Pacific Journal of Information Systems,* vol. 19, no. 4, pp. 77–100, 2009.

[3] M. Aslam, "User Resistance in Post Erp Implementation," *Journal of Business Process Management,* vol. 17, pp. 266-275, 2011.

[4] T. H. Baek and M. Kim, "Is ChatGPT scary good? How user motivations affect creepiness and trust in generative artificial intelligence," *Telematics and Informatics,* vol. 83, 2023, doi: https://doi.org/10.1016/j.tele.2023.102030.

[5] A. Dimoka, P. Pavlou, and F. D. Davis, "NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research," *Information Systems Research,* vol. 22, no. 4, pp. 687-702, Dec 2011.

[6] M. Riedl and P.-M. Léger, *Fundamentals of NeuroIS. Information Systems and the Brain*. Springer, 2016.

[7] R. Riedl *et al.*, " On the foundations of NeuroIS: Reflections on the Gmunden Retreat," *Communications of the Association for Information Systems,* vol. 27, pp. 243–264, 2010.

[8] R. Riedl and P.-M. Léger, *Fundamentals of NeuroIS: Information systems and the brain*. Springer, 2016.

[9] F. Marimon, M. Mas-Machuca, and A. Akhmedova, "Trusting in Generative AI: Catalyst for Employee Performance and Engagement in the Workplace," *International Journal of Human-Computer Interaction,* vol. 41, no. 11, pp. 7076–7091, 2025, doi: https://doi.org/10.1080/10447318.2024.2388482.

[10] H. Zhang, L. Zhu, A. Zhang, and K. Shohruh, "The influence of generative artificial intelligence usage on employees' innovative job performance," *PloS one,* vol. 21, no. 1, 2026, doi: https://doi.org/10.1371/journal.pone.0327786.

[11] Y. Zhang, "GPT is all you need," *Frontiers of Psychology* vol. 10, no. 16, 2025 doi: 10.3389/fpsyg.2025.1549755.

[12] A. Ghodmare, S. Bobade, and K. Dua, *Integrating Sales and Distribution in SAP S/4HANA*. Rheinwerk Publishing, 2023, p. 1493223291.

[13] C. Li and A. K. Mutlu, " The Interplay of Personal Learning Experience, Social Emotional Learning and Autonomy in Promoting Well-Being in AI-Assisted Language Learning: A Self-Determination Theory Approach," *European Journal of Education,* vol. 60, no. 4, 2025, doi: https://doi-org.ezp.slu.edu/10.1111/ejed.70354.

[14] J. Sweller, "The development of cognitive load theory: Replication crises and incorporation of other theories can lead to theory expansion," *Educational Psychology Review,* vol. 35, no. 3, pp. 1-25, 2023.

[15] R. E. Mayer, "The past, present, and future of the cognitive theory of multimedia learning," *Educational Psychology Review,* vol. 36, no. 1, pp. 1-25, 2024.

[16] L. Fiorella, "Making sense of generative learning," *Educational Psychology Review,* vol. 35, no. 50, pp. 1-50, 2023.

[17] S. Kuldas, S. Lata, N. I. Hairul, and H. Shahabuddin, "Greater Cognitive Effort for Better Learning: Tailoring an Instructional Design for Learners with Different Levels of Knowledge and Motivation," *Psychologica Belgica,* vol. 54, no. 4, pp. 350-373, 2014.

[18] M. C. Wittrock, "Generative processes of comprehension," *Educational Psychologist,* vol. 24, no. 4, pp. 345–376, 1989.

[19] L. Fiorella and R. E. Mayer, *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press, 2015.

[20] A. Dimoka *et al.*, "On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS," *MIS Quarterly,* vol. 36, no. 3, pp. 679–702, 2012.

[21] A. T. Duchowski, "A Breadth-First Survey of Eye Tracking Applications," *Behavior Research Methods, Instruments, & Computers,* vol. 34, no. 4, pp. 455-470, 2002.

[22] H. J. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: Methods and constructs.," *International Journal of Industrial Ergonomics,* vol. 24, pp. 631-645, 1999.

[23] A. Poole and L. J. Ball, "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects," in *Encyclopedia of Human Computer Interaction*, C. Ghaoui Ed. Hershey, PA: Idea Group, 2006, pp. 211-219.

Stacking CTML Principles in Video Lessons: A Multimodal Feasibility Case-Series (Eye, Facial, GSR) and Early Steps Toward a Neuroadaptive Learning System

Craig Chin[1], Ordene Edwards[1], May Bantan[2], Armando Barreto[3], Andrea Taylor[1] and Adriane Randolph[1]

[1] Kennesaw State University, Kennesaw, GA, USA
{cchin2,oedward4,atayl251,arandol3}@kennesaw.edu
[2] University of Colorado Colorado Springs, Colorado Springs, CO, USA
mbantan@uccs.edu
[3] Florida International University, Miami, FL, USA
barretoa@fiu.edu

Abstract. This feasibility case-series examines whether stacking principles from the Cognitive Theory of Multimedia Learning (CTML) produces measurable multimodal engagement signatures during video-based instruction and whether richer support relates to more coherent engagement profiles. We contrast a CTML-rich High-Engagement (HE) video with a Low-Engagement (LE) version ($N = 7$; $HE = 4$, $LE = 3$) while recording synchronized eye-tracking, webcam-based facial expression indicators, and electrodermal activity (EDA). Participant-level summary metrics are compared using descriptive statistics, Spearman correlation heat maps, and nonparametric Mann–Whitney U tests with Cliff's δ effect sizes. Results suggest directional differences between HE and LE in affective and autonomic indicators and partial cross-modal structure within conditions, though with substantial overlap and variability. Findings are interpreted as feasibility evidence that CTML-aligned instructional manipulation can shape the organization of multimodal engagement signals, motivating larger-scale and time-resolved modeling in future neuroadaptive learning research.

Keywords: Cognitive Online Learning · Cognitive Theory of Multimedia Learning (CTML) · eye tracking · electrodermal activity · engagement recognition

Introduction

Online video instruction has become a dominant mode of undergraduate learning, particularly in STEM disciplines where scalability and flexibility are essential. Yet sustained engagement during video learning remains fragile: learners frequently experience attentional lapses, fluctuating cognitive load, and uneven regulation of effort, all of which can undermine comprehension and transfer [1], [2], [3]. While prior work has established that engagement matters for learning outcomes, much of the existing evidence relies on coarse proxies (e.g., self-report or post-hoc performance) or correlational associations between behavioral or physiological signals and learning [1], [4]. As a result, it remains difficult to specify when, how, and why instructional design choices alter learner engagement in ways that are actionable for neuroadaptive systems.

Recent advances in multimodal sensing, eye tracking, facial expression analysis, and autonomic physiology such as electrodermal activity (EDA), have enabled increasingly fine-grained measurement of learner state during video learning [3], [4], [5], [6]. However, a persistent limitation is that engagement signals are often modeled independently of the instructional manipulations that are presumed to influence them. Gaze dispersion, facial expressiveness, and arousal are treated as indicators of engagement, but the causal relationship between instructional design features and these signals is rarely tested [2], [4]. Without such causal anchors, engagement detection risks becoming an exercise in pattern recognition rather than a mechanism-informed basis for adaptive intervention [1].

The Cognitive Theory of Multimedia Learning (CTML) offers a principled framework for addressing this gap. CTML specifies how instructional design features influence learning through well-articulated cognitive mechanisms: dual-channel processing, limited working memory capacity, and active selection, organization, and integration of information [7], [8]. Design principles such as signaling, coherence, modality, and segmentation are intended to reduce extraneous processing, manage essential cognitive load, and foster generative processing [7], [8], [9], [10]. Importantly, these principles are not merely design variations; they function as interventions that target specific attentional and cognitive control processes during instruction. This makes CTML-aligned manipulations particularly well-suited as causal levers for studying engagement dynamics.

Despite extensive outcome-focused evidence supporting CTML principles [9], [10], relatively little work has examined how CTML-aligned instructional support manifests in time-resolved, multimodal engagement signals. Reviews of eye-tracking research on instructional video caution that mappings from gaze metrics to cognition are often underspecified and call for multimodal measurement aligned to theory [2]. In educational psychophysiology, EDA is widely used as an index of arousal and effort but exhibits methodological variability and interpretive ambiguity across studies [11], while related autonomic measures such as HRV require cautious interpretation for attention and cognitive load [12]. Multimodal engagement studies increasingly combine gaze, facial, and physiological signals [4], [13], yet these signals are typically analyzed without explicit linkage to instructional events or principled design contrasts grounded in learning theory.

The present work addresses this gap by treating CTML-aligned instructional design as a causal manipulation and examining its relationship to multimodal engagement signals. Using a small feasibility case series, we contrast a High-Engagement (HE) video that stacks multiple CTML principles with a Low-Engagement (LE) comparison video that applies principles more sparingly. We focus on three complementary modalities, eye tracking, facial expression indicators, and EDA, selected because they index attention allocation, affective expressiveness, and autonomic arousal/regulation, respectively [3], [5], [6], [11]. Rather than asserting definitive effects, our goal is to investigate whether theoretically motivated instructional differences are reflected in structured, convergent patterns across modalities. Accordingly, we pose two research questions.

- RQ1 (Cross-modal convergence). To what extent do gaze, facial expression, and electrodermal measures exhibit convergent patterns consistent with a shared learner engagement construct during video-based learning?
- RQ2 (Exposure–response). Does CTML-rich instructional design relate to differences in the organization or efficiency of multimodal engagement signals compared to lower levels of CTML support?

By framing CTML principles as experimentally manipulable causes rather than background design heuristics, this study takes an initial step toward grounding neuroadaptive engagement modeling in learning theory. The results are intended to inform subsequent large-scale, time-resolved analyses and the development of interpretable, safety-aware neuroadaptive learning systems [1].

Related Work

Neuroadaptive / Affect-Adaptive Learning Systems. A growing body of research on neuroadaptive (or more broadly affect- and state-adaptive) learning systems conceptualize engagement as a time-bounded, context-dependent, multicomponent construct encompassing cognitive, affective, and behavioral processes [1]. In this framing, engagement detection refers to the real-time estimation of short-timescale learner states (e.g., attentional lapses, mind wandering, disengagement intensity, or affective arousal) from observable signals, with the goal of enabling targeted instructional responses that support learning persistence, comprehension, and transfer [1].

Across learning technologies, one prominent detection target is mind wandering, a specific form of attentional disengagement that occurs frequently during technology-mediated learning and is associated with reduced learning outcomes [3]. Another common target is graded engagement intensity, often modeled as ordinal or continuous rather than binary (e.g., disengaged to highly engaged), particularly in video-based and online learning contexts [5], [6]. Recent work further emphasizes the importance of label quality and intra-class variability, proposing engagement annotations aligned with psychological theory rather than ad hoc heuristics [14].

In terms of sensing modalities, many neuroadaptive systems rely on eye tracking as a primary channel for attention estimation, as gaze provides temporally precise indicators of visual attention allocation and attentional lapses [3]. Classroom and in-the-wild studies demonstrate that commercial off-the-shelf eye trackers can support student-independent mind-wandering detection and limited real-time deployment [3]. Moving beyond detection, attention-aware systems have used gaze-based signals to trigger re-engagement strategies (e.g., prompting, pacing adjustments), with benefits often moderated by learner characteristics such as prior knowledge or baseline engagement [3].

A complementary line of work focuses on adaptive reading and learning interfaces that detect attentional disengagement and deliver event-contingent interventions. The

Eye-Mind Reader system exemplifies this approach by using gaze-based mind-wandering detection to trigger self-explanation and rereading prompts, resulting in improved delayed comprehension relative to yoked controls [15]. These systems illustrate a common neuroadaptive pipeline: (i) learner-state estimation, (ii) micro-intervention, and (iii) evaluation using immediate and delayed learning outcomes, consistent with proactive and reactive engagement enhancement frameworks [1].

Operationally, engagement detection spans a wide methodological range, from short-clip facial expression classification [5], [6], to sequence models over time-series features in online environments [16], to sensor-free affect detection based on interaction logs for large-scale deployment [17]. More recent frameworks advocate multimodal fusion, combining visual, audio, physiological, and behavioral streams, and integrating closed-loop adaptive feedback in higher education settings [18]. However, reviews consistently emphasize that engagement is difficult to define and measure, and that adaptive systems must be evaluated with attention to validity, interpretability, fairness, and context sensitivity [1], [4].

Multimodal Engagement Detection in Video Learning (Gaze, Facial Expression, EDA). Multimodal engagement detection in video-based and online learning environments typically integrates behavioral attention proxies (e.g., gaze, head pose, on-screen actions), affective proxies (e.g., facial expressions or action units), and autonomic physiology (e.g., electrodermal activity, heart rate, or heart-rate variability), with the aim of improving robustness beyond any single channel [4], [13]. This approach reflects the recognition that engagement is not directly observable and that individual modalities capture complementary aspects of learner state.

Early work established that student engagement can be inferred from facial expressions and that automated classifiers can approach human judgments for coarse engagement distinctions, while finer-grained, multi-level classification remains challenging [5]. Building on this foundation, Savchenko et al. demonstrated that a single, efficient facial expression recognition network can simultaneously estimate emotions and engagement levels in online learning video, enabling real-time processing and group-level analysis [6]. Facial indicators have also been used as proxies for cognitive engagement in problem-solving contexts. Li, Lajoie, and colleagues showed that facial behaviors predict cognitive engagement states and differentiate deeper versus more superficial learning strategies in intelligent tutoring environments [19]. In game-based learning, Ninaus et al. linked facial emotion dynamics to emotionally engaging instructional contexts using machine-learning models [20]. More recent deep spatiotemporal models for e-learning video (e.g., EfficientNetV2-L combined with recurrent architectures) reflect continued methodological advances, though reported accuracy remains sensitive to dataset constraints and labeling quality [16].

Eye tracking provides a comparatively direct window into visual attention allocation, supporting fine-grained modeling of attentional stability and lapses during learning. Hutt et al. demonstrated that gaze features collected in authentic classroom settings can support student-independent mind-wandering detection and generalize across learners when appropriate validation strategies are used [3]. In adaptive contexts, gaze-based detection has been successfully coupled with interventions. Mills et al.'s Eye-Mind Reader uses gaze to identify mind wandering during reading and triggers self-explanation and rereading prompts, yielding gains in delayed comprehension [15]. At the opposite end of the spectrum, Hutt et al.'s "Time to Scale" illustrates how engagement-related affect can be inferred without sensors using platform interaction traces, enabling deployment across tens of thousands of learners but at the cost of reduced temporal precision [17].

A canonical multimodal strategy is to explicitly align engagement facets with modality channels (affective, behavioral, and cognitive) and to fuse these channels during inference. Yue et al. combined facial expressions, eye movements, and interaction data to recognize multidimensional engagement and learning performance, while also addressing the reliability of self-report annotations [13]. Chang et al. proposed an ensemble model integrating face and body tracking features (e.g., OpenFace and OpenPose outputs), reflecting common practice in combining facial, gaze, head pose, and posture cues [21]. More recently, Kumar et al. extended multimodal fusion by integrating visual and audio streams using cross-modality pipelines and transformer-based encoders, highlighting the contribution of audio features in engagement prediction [22].

Recent work has emphasized the need for higher-quality datasets and better coverage of online learning behaviors. The CMOSE dataset was explicitly designed for online student engagement, providing multimodal visual and audio data with psychologically informed, high-quality labels and demonstrating the complexity of ordinal engagement classification [14]. Vision-centric pipelines for online classes increasingly incorporate action recognition to distinguish disengagement from legitimate off-screen activities (e.g., note-taking) [23]. Other approaches extend multimodality with non-contact physiology, such as remote photoplethysmography (rPPG), showing that cardiac features can complement behavioral cues in online meetings and learning-adjacent settings [24], [25].

Within multimodal engagement research, electrodermal activity (EDA/GSR) is commonly used as an index of autonomic arousal and effort and is often positioned as complementary to gaze and facial indicators [4]. Reviews emphasize that EDA interpretation in learning contexts requires attention to signal dynamics and morphology rather than reliance on global averages [26]. Multimodal neurophysiological studies further demonstrate that different educational materials elicit distinct patterns of engagement and cognitive processing when assessed using combinations of EEG, EDA, and cardiovascular measures [27]. In e-learning contexts, multimodal physiological fusion has been applied to emotion recognition, providing an affective component that can inform multicomponent engagement models [28].

Taken together, this literature converges on a pragmatic view: gaze and facial cues offer interpretable, temporally precise indicators of attention and affect; physiological signals add sensitivity to effort and regulation; and audio or interaction traces improve scalability. However, most multimodal engagement studies focus on detecting engagement states rather than examining how theoretically grounded instructional manipulations shape these signals over time [1], [4], [14]. This gap motivates the present work's emphasis on CTML-aligned design contrasts as a means of grounding multimodal engagement analysis in learning theory.

CTML, Engagement Signals, and an Unresolved Gap. Prior research provides partial links between CTML-aligned instructional design and engagement-related signals, but these links remain fragmented. Studies of signaling and cueing show that visual cues reliably redirect learners' gaze toward task-relevant information, consistent with CTML's attentional selection mechanism, and eye-tracking reviews confirm that cueing alters fixation patterns during video learning [2], [9], [29]. Segmentation and self-explanation, principles aimed at managing cognitive load and fostering generative processing, are robustly associated with improved learning outcomes and reduced subjective load [10], [11], [12]. Separately, psychophysiological and multimodal studies demonstrate that instructional materials elicit distinct patterns of arousal and engagement as indexed by EDA, cardiovascular measures, and facial expressions [11], [27], [28]. However, CTML studies rarely examine continuous multimodal engagement signals, while multimodal engagement detection studies seldom manipulate CTML principles as causal variables [1], [4], [13]. Consequently, it remains unclear whether CTML-aligned support produces coherent, convergent changes across independent engagement modalities during learning.

Methods

Recruitment. The participants (to date: 4 HE, 3 LE) for this study were taken from the undergraduate student population of a university institution. The inclusion criteria for this study are that each participant must be 18 or older and they are not currently enrolled in or have previously been enrolled in Physics II and Circuit Analysis I courses. These two courses review content that overlaps with the topic of the learning videos used in this study.

Learning Video Design. The topic reviewed in the learning videos for this study is DC series circuit analysis. This topic was delivered using two versions: a CTML-rich High-Engagement (HE) version and a comparison Low-Engagement (LE) version. The HE version employed personalization/voice, signaling, coherence (text minimization), temporal/spatial contiguity, and multimedia/modality where appropriate, while the LE version limited principle use to pedagogical necessity (e.g., accurate schematics).

Apparatus. Multimodal data were captured using the iMotions 11 software application, which provides an integrated research platform for synchronized, time-aligned collection, visualization, and export of multimodal data all aligned to stimulus timelines to support unified multimodal engagement analysis [30]. The input modalities for this study are screen-based eye tracking via the EyeTech VT3 Mini, webcam-based facial expression analysis using Affectiva/Affdex software and GSR/EDA signals using the Shimmer3 GSR+ unit (Fig. 1).

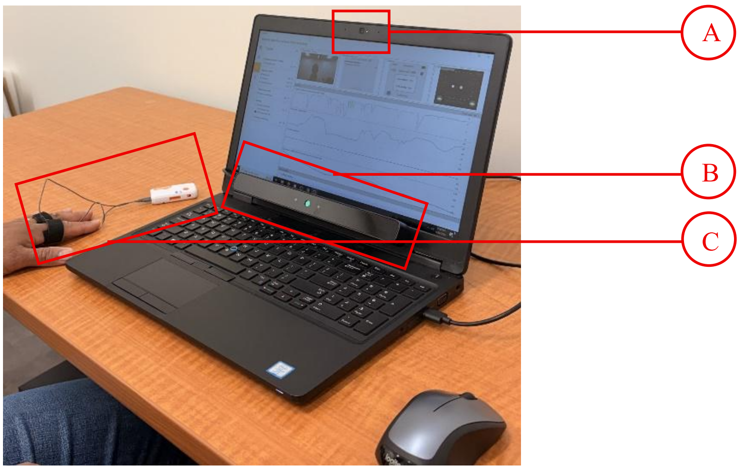


Fig. 17. Experiment Apparatus Set Up (A – Webcam, B – EyeTech VT3 Mini Eye Tracker, C – Shimmer3 GSR+ unit)

Surveys. Pre/post self-report instruments adapted items from the Motivated Strategies for Learning Questionnaire (MSLQ) [31] and the Student Engagement in Schools Questionnaire (SESQ) [32] to gauge cognitive, affective, and behavioral engagement. The items adapted from MSLQ were used to create 4 subscales that measured cognitive and metacognitive strategies: elaboration, critical thinking, metacognitive self-regulation (SR), and effort self-regulation. The SESQ items formed a subscale that measured affective engagement.

Multimodal Metrics. The following multimodal metrics were employed to analyze learner engagement as they viewed the learning videos:

i. Eye tracking: gaze dispersion (e.g., average pairwise fixation distance).
ii. EDA: phasic SCR rate (# responses/min) and mean SCR amplitude across the lesson; tonic SCL mean and linear slope.
iii. Facial indicators: mean Affectiva engagement (expressiveness) and mean valence.

Analysis Plan. Given the small sample (N=7), we compute participant-level means over the full viewing period for each multimodal metric and conduct nonparametric HE–LE contrasts using two-sided Mann–Whitney U with effect sizes (Cliff's δ) to address RQ2 (exposure–response). To address RQ1 (cross-modal convergence), we estimate Spearman correlation matrices among gaze, facial, EDA measures, and survey Δ subscales at the participant level.

Results and Discussion

Given the small sample ($N = 7$; $HE = 4$, $LE = 3$), results are reported descriptively with emphasis on patterns, dispersion, and directional tendencies rather than statistical significance. Analyses address RQ1 (cross-modal convergence) and RQ2 (exposure–response), with conclusions framed as exploratory.

Cross-modal convergence (RQ1). Cross-modal convergence was examined using participant-level Spearman correlation heat maps and summary-level descriptives. Figures 2 and 3 present the correlation structures for the HE and LE groups, respectively, showing how gaze, facial expression, EDA, and survey change scores covary within each condition.

In the HE group (Fig. 2), partial alignment is visible among affective and autonomic indicators. Facial engagement tends to covary positively with SCR rate, phasic amplitude, tonic skin conductance level (SCL) and SCL slope, while Facial valence tends covary negatively with these EDA measures. Also, phasic EDA measures cluster together. This pattern suggests localized cross-modal structure, consistent with subsets of engagement-related processes moving together under CTML-rich instructional support. However, associations between behavioral or physiological signals and self-report deltas are weak and inconsistent.

In contrast, the LE group (Fig. 3) exhibits more fragmented and mixed-direction relationships, including sign reversals between facial and tonic EDA measures and weaker clustering overall. Survey change scores again show little systematic alignment with physiological or behavioral metrics.

Overall, the heat maps indicate that multimodal signals can form partially coherent structures and that these structures differ by instructional condition. At the same time, given the small N and instability of correlation estimates, the evidence does not support strong claims of robust cross-modal convergence tracking subjective engagement change. Accordingly, RQ1 is interpreted as providing feasibility and directional differentiation, rather than construct validation.

Exposure–response to CTML support (RQ2). To assess exposure–response, participant-level summary metrics were contrasted between HE and LE. Table 2 reports medians along with Mann–Whitney U tests and Cliff's δ effect sizes for behavioral and physiological measures. Differences are modest and highly overlapping, as expected with limited power, but several directional tendencies emerge.

Median facial engagement, valence, and tonic EDA indicators (mean SCL and SCL slope) tend to be higher in HE than LE (Table 2), consistent with sustained arousal or regulatory effort during the CTML-rich video. In contrast, gaze dispersion and phasic EDA measures show substantial overlap across conditions. Survey results (Table 1) are mixed: affective engagement change favors HE, whereas elaboration and critical-thinking deltas are not uniformly higher and in some cases favor LE.

Cliff's δ values across Tables 1 and 2 indicate small-to-moderate directional effects rather than clear separation. Inspection of within-condition dispersion further suggests that HE participants exhibit more compact engagement profiles on facial and tonic EDA measures, whereas LE participants display greater variability across channels.

Summary. In summary, the results demonstrate that multimodal engagement data can be captured and compared at the participant level, and that CTML-rich instruction is associated with directional and organizational differences in engagement signals (Figs. 2–3; Tables 1–2). However, substantial overlap and reverse tendencies underscore the need for larger samples and time-resolved analyses. The findings are therefore best interpreted as feasibility evidence motivating continued investigation rather than definitive tests of RQ1 or RQ2.

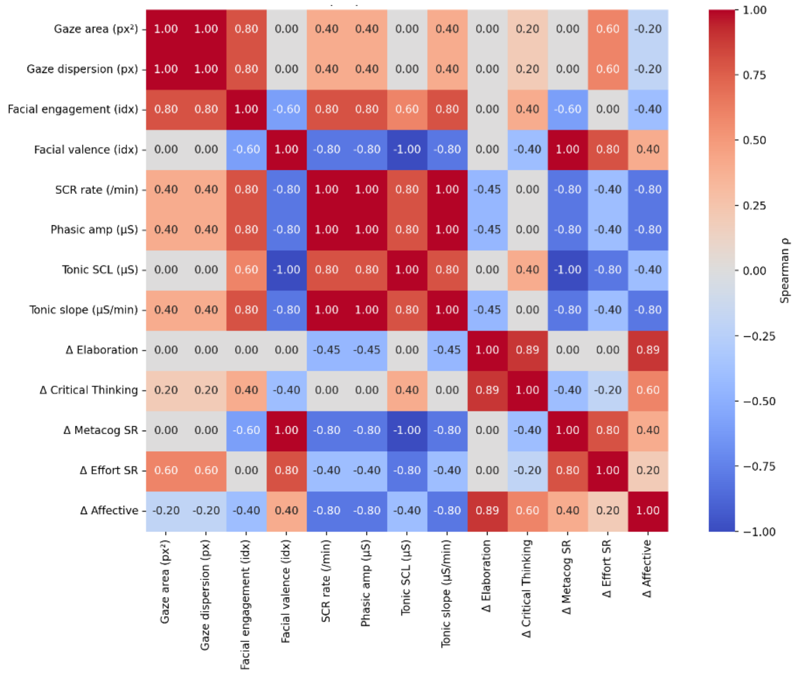


Fig. 2. Spearman Correlation Heatmap for High Engagement (HE) Group

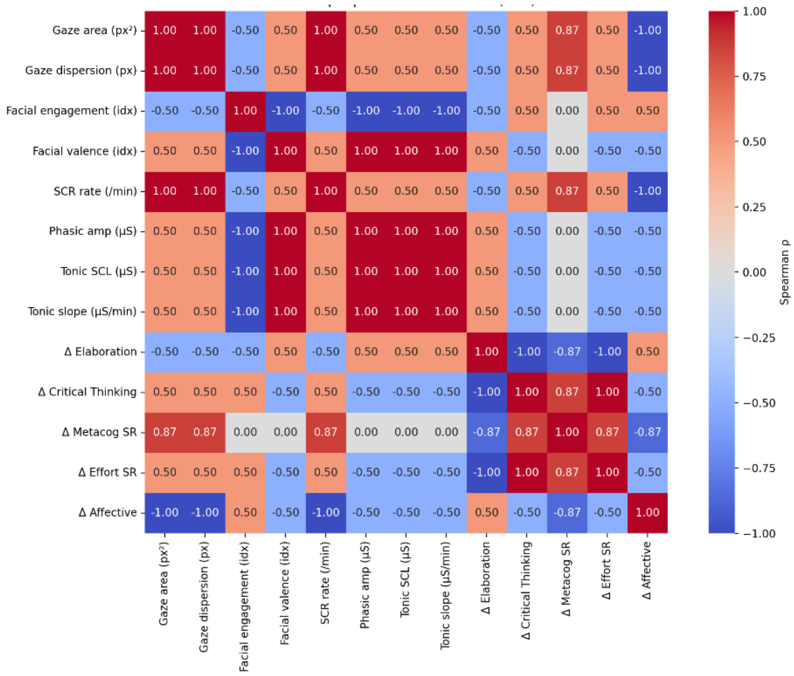


Fig. 3. Spearman Correlation Heatmap for Low Engagement (LE) Group

Table 11. Between-group (HE vs. LE) differences in survey change scores (Δ = Post − Pre). Medians shown; Mann–Whitney U with Cliff's δ

Subscale	n_HE	n_LE	Median Δ HE	Median Δ LE	U	p (expl.)	Cliff's δ
Elaboration	4	3	0.000	1.250	4.000	0.589	-0.333
Critical Thinking	4	3	-0.500	0.000	7.500	0.721	0.250
Metacognitive SR	4	3	0.875	0.750	5.500	1.000	-0.083
Effort SR	4	3	0.875	0.750	6.000	1.000	0.000
Affective Engagement	4	3	0.300	0.000	10.000	0.229	0.667

Table 2. Between-group (HE vs. LE) differences in behavioral/physiological signals. Medians shown; Mann–Whitney U with Cliff's δ

Variable (mean values)	n_HE	n_LE	Median HE	Median LE	U	p	Cliff's δ
Gaze dispersion bounding box (px^2)	4	3	5.72E+05	4.86E+05	8	0.63	0.333
Gaze dispersion (mpd) (px)	4	3	4.08E+02	4.37E+02	5	0.86	-0.167
FEA Engagement	4	3	6.37E+00	4.63E+00	8	0.63	0.333
FEA Valence	4	3	4.72E-01	-1.50E+00	8	0.63	0.333
SCR rate per min	4	3	3.92E+00	5.04E+00	7	0.86	0.167
SCR Amp (uS)	4	3	6.32E-02	5.38E-02	8	0.63	0.333
Tonic SCL (uS)	4	3	4.37E+00	3.76E+00	8	0.63	0.333
Tonic SCL slope (uS/min)	4	3	4.93E-02	3.59E-02	7	0.86	0.167

Conclusion and Future Work

This study demonstrates the feasibility of capturing and comparing multimodal engagement signals during CTML-aligned video learning. Directional differences and profile-level organization across gaze, facial expression, and EDA suggest that stacked CTML principles may shape engagement dynamics, though results remain provisional given the small sample and summary-level analyses. Future work will expand sample size, incorporate time-resolved and event-aligned modeling, and examine how specific CTML principles drive moment-to-moment engagement changes. These steps are necessary to establish robust causal links and inform the design of interpretable, responsible neuroadaptive learning systems.

References

[1] B. M. Booth, N. Bosch, and S. K. D'Mello, "Engagement Detection and Its Applications in Learning: A Tutorial and Selective Review," *Proceedings of the IEEE*, vol. 111, no. 10, pp. 1398–1422, Oct. 2023, doi: 10.1109/JPROC.2023.3309560.

[2] R. Deng and Y. Gao, "A review of eye tracking research on video-based learning," *Education and Information Technologies 2022 28:6*, vol. 28, no. 6, pp. 7671–7702, Dec. 2022, doi: 10.1007/s10639-022-11486-7.

[3] S. Hutt *et al.*, "Automated gaze-based mind wandering detection during computerized learning in classrooms," *User Modeling and User-Adapted Interaction 2019 29:4*, vol. 29, no. 4, pp. 821–867, Jun. 2019, doi: 10.1007/S11257-019-09228-5.

[4] A. Becerra, R. Cobos, and C. Lang, "Enhancing online learning by integrating biosensors and multimodal learning analytics for detecting and predicting student behaviour: a review," *Taylor & FrancisA Becerra, R Cobos, C LangBehaviour & Information Technology, 2025•Taylor & Francis*, 2025, doi: 10.1080/0144929X.2025.2562322.

[5] J. Whitehill, Z. Serpell, Y. Lin, … A. F.-I. T. on, and undefined 2014, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *ieeexplore.ieee.orgJ Whitehill, Z Serpell, YC Lin, A Foster, JR MovellanIEEE Transactions on Affective Computing, 2014•ieeexplore.ieee.org*, Accessed: Apr. 22, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6786307/

[6] A. Savchenko, … L. S.-I. transactions on, and undefined 2022, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *ieeexplore.ieee.orgAV Savchenko, LV Savchenko, I MakarovIEEE transactions on affective computing, 2022•ieeexplore.ieee.org*, Accessed: Apr. 22, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9815154/

[7] R. E. Mayer, "The Past, Present, and Future of the Cognitive Theory of Multimedia Learning," *Educational Psychology Review 2024 36:1*, vol. 36, no. 1, pp. 8-, Jan. 2024, doi: 10.1007/S10648-023-09842-1.

[8] R. E. . Mayer and Logan. Fiorella, *The Cambridge handbook of multimedia learning*. Cambridge University Press, 2022.

[9] D. Alpizar, O. O. Adesope, and R. M. Wong, "A meta-analysis of signaling principle in multimedia learning environments," *Educational Technology Research and Development 2020 68:5*, vol. 68, no. 5, pp. 2095–2119, Feb. 2020, doi: 10.1007/s11423-020-09748-7.

[10] G. D. Rey, M. Beege, S. Nebel, M. Wirzberger, T. H. Schmitt, and S. Schneider, "A Meta-analysis of the Segmenting Effect," *Educational Psychology Review 2019 31:2*, vol. 31, no. 2, pp. 389–419, Jan. 2019, doi: 10.1007/s10648-018-9456-4.

[11] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, and I. Molenaar, "Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review," *Sensors 2021, Vol. 21, Page 7869*, vol. 21, no. 23, p. 7869, Nov. 2021, doi: 10.3390/s21237869.

[12] H. J. Kim, Y. Park, and J. Lee, "The validity of heart rate variability (HRV) in educational research and a synthesis of recommendations," *SpringerHJ Kim, Y Park, J LeeEducational Psychology Review, 2024•Springer*, vol. 36, no. 2, p. 42, Jun. 2024, doi: 10.1007/s10648-024-09878-x.

[13] J. Yue *et al.*, "Recognizing multidimensional engagement of e-learners based on multi-channel data in e-learning environment," *ieeexplore.ieee.orgJ Yue, F Tian, KM Chao, N Shah, L Li, Y Chen, Q ZhengIEEE Access, 2019•ieeexplore.ieee.org*, Accessed: Apr. 22, 2026. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8866724/

[14] C.-H. Wu *et al.*, "CMOSE: comprehensive multi-modality online student engagement dataset with high-quality labels," *openaccess.thecvf.comCH Wu, S Liu, X Huang, X Wang, R Zhang, L Minciullo, WK Yiu, K Kwan, KT ChengProceedings of the IEEE/CVF Conference on Computer Vision and, 2024•openaccess.thecvf.com*, Accessed: Apr. 22, 2026. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024W/ABAW/html/Wu_CMOSE_Comprehensive_Multi-Modality_Online_Student_Engagement_Dataset_with_High-Quality_Labels_CVPRW_2024_paper.html

[15] C. Mills, J. Gregg, R. Bixler, and S. K. D'Mello, "Eye-mind reader: An intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering," *Taylor & FrancisC Mills, J Gregg, R Bixler, SK D'MelloHuman–Computer Interaction, 2021•Taylor & Francis*, vol. 36, no. 4, pp. 306–332, 2021, doi: 10.1080/07370024.2020.1716762.

[16] F. M. Shiri, E. Ahmadi, M. Rezaee, and T. Perumal, "Detection of student engagement in E-learning environments using EfficientNetV2-L together with RNN-based models," *psasir.upm.edu.myF Shiri, E Ahmadi, M Rezaee, T PerumalJournal of Artificial Intelligence, 2024•psasir.upm.edu.my*, doi: 10.32604/jai.2024.048911.

[17] S. Hutt, J. F. Grafsgaard, and S. K. D'Mello, "Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year," *dl.acm.orgS Hutt, JF Grafsgaard, SK D'MelloProceedings of the 2019 CHI conference on human factors in computing systems, 2019•dl.acm.org*, May 2019, doi: 10.1145/3290605.3300726.

[18] Z. Dan, Z. Yali, and Z. Tong, "A multimodal AI framework for real-time student engagement detection and adaptive feedback in higher education," *Journal of King Saud University Computer and Information Sciences 2026*, Mar. 2026, doi: 10.1007/S44443-026-00639-0.

[19] S. Li, S. Lajoie, J. Zheng, H. Wu, H. C.-C. & Education, and undefined 2021, "Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving," *ElsevierS Li, SP Lajoie, J Zheng, H Wu, H ChengComputers & Education, 2021•Elsevier*, Accessed: Apr. 22, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131520303122

[20] M. Ninaus, S. Greipl, K. Kiili, A. Lindstedt, … S. H.-C. &, and undefined 2019, "Increased emotional engagement in game-based learning–A machine learning approach on facial emotion detection data," *ElsevierM Ninaus, S Greipl, K Kiili, A Lindstedt, S Huber, E Klein, HO Karnath, K MoellerComputers & Education, 2019•Elsevier*, Accessed: Apr. 22, 2026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131519301940

[21] C. Chang, L. Chen, C. Zhang, and Y. Liu, "An ensemble model using face and body tracking for engagement detection," *dl.acm.orgC Chang, C Zhang, L Chen, Y LiuProceedings of the 20th ACM international conference on multimodal interaction, 2018•dl.acm.org*, pp. 616–622, Oct. 2018, doi: 10.1145/3242969.3264986.

[22] D. Kumar, S. Madan, P. Singh, A. Dhall, and B. Raman, “Towards engagement prediction: a cross-modality dual-pipeline approach using visual and audio features,” *dl.acm.orgD Kumar, S Madan, P Singh, A Dhall, B RamanProceedings of the 32nd ACM International Conference on Multimedia, 2024•dl.acm.org*, pp. 11383–11389, Oct. 2024, doi: 10.1145/3664647.3688986.

[23] Y. Qi, L. Zhuang, H. Chen, X. Han, and A. Liang, “Evaluation of Students’ Learning Engagement in Online Classes Based on Multimodal Vision Perspective,” *Electronics 2024, Vol. 13, Page 149*, vol. 13, no. 1, p. 149, Dec. 2023, doi: 10.3390/ELECTRONICS13010149.

[24] A. Singh, K. Goyal, N. Verma, P. Kumar, X. Li, and A. Singh, “VisioPhysioENet: Visual Physiological Engagement Detection Network,” Aug. 2025, Accessed: Apr. 22, 2026. [Online]. Available: https://arxiv.org/pdf/2409.16126v2

[25] A. Vedernikov, Z. Sun, V.-L. Kykyri, M. Pohjola, M. Nokia, and X. Li, “Analyzing Participants’ Engagement during Online Meetings Using Unsupervised Remote Photoplethysmography with Behavioral Features,” *openaccess.thecvf.comA Vedernikov, Z Sun, VL Kykyri, M Pohjola, M Nokia, X LiProceedings of the IEEE/CVF Conference on Computer Vision and, 2024•openaccess.thecvf.com*, Accessed: Apr. 22, 2026. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024W/CVPM/html/Vedernikov_Analyzing_Participants_Engagement_during_Online_Meetings_Using_Unsupervised_Remote_Photoplethysmography_CVPRW_2024_paper.html

[26] K. Nandipati, S. Pal, R. M.- Sensors, and undefined 2024, “Electrodermal activity (EDA) morphologies and prediction of engagement with simple moving average crossover: a mixed-method study,” *mdpi.comKK Nandipati, S Pal, R MitraSensors, 2024•mdpi.com*, Accessed: Apr. 22, 2026. [Online]. Available: https://www.mdpi.com/1424-8220/24/14/4565

[27] V. Ronca, P. Aricò, L. Tamborra, A. Biagi, and G. Di Flumeri, “A Multimodal Neurophysiological Approach to Evaluate Educational Contents in Terms of Cognitive Processes and Engagement,” *Bioengineering 2025, Vol. 12, Page 597*, vol. 12, no. 6, p. 597, May 2025, doi: 10.3390/BIOENGINEERING12060597.

[28] S. P. Cherian, “E-learning emotion recognition by fusing ECG, EEG and EDA signals based on Bayes optimal cue integration model,” *Multimedia Tools and Applications 2026 85:3*, vol. 85, no. 3, pp. 218-, Feb. 2026, doi: 10.1007/S11042-026-21152-6.

[29] P. Mautone, R. M.-J. of educational Psychology, and undefined 2001, “Signaling as a cognitive guide in multimedia learning.,” *psycnet.apa.org*, Accessed: Jan. 30, 2026. [Online]. Available: https://psycnet.apa.org/fulltext/2001-06601-012.html

[30] “iMotions | Powering Human Insights - Biometric Research.” Accessed: Mar. 06, 2026. [Online]. Available: https://imotions.com/?srsltid=AfmBOorBrwWq_nHFXI3ResB1Yn93Poa3TY5LSJPO_vaQHePJidQN_per

[31] P. R. Pintrich and E. V. De Groot, “Motivated Strategies for Learning Questionnaire,” *PsycTESTS Dataset*, Mar. 2012, doi: 10.1037/t09161-000.

[32] S. Hart, K. Stewart, and S. J.-C. S. Psychology, “Student Engagement in Schools Questionnaire,” *psycnet.apa.orgSR Hart, K Stewart, SR JimersonContemporary School Psychology•psycnet.apa.org*, Accessed: Mar. 06, 2026. [Online]. Available: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft43831-000

What Shall Artificial Intelligence Recognize When Programmed for the Purpose of Emotion Recognition?

Peter Walla[1,2,3*]

[1] Faculty of Psychology, Freud CanBeLab, Sigmund Freud Private University, Freudplatz 1, 1020 Vienna, Austria
[2] Faculty of Medicine, Sigmund Freud Private University, Freudplatz 3, 1020 Vienna, Austria
[3] School of Psychology, Centre for Translational Neuroscience and Mental Health Research, University of Newcastle, University Drive, Callaghan, NSW, Australia
peter.walla@sfu.ac.at

Abstract. The rapid advancement of Artificial Intelligence (AI) systems designed for "emotion" recognition purposes has reached a critical impasse, primarily due to the lack of a unified definition of "emotion." Current models are built upon inconsistent frameworks that conflate physiological responses, subjective experiences, cognitive appraisals, and expressive displays. This conceptual ambiguity hinders communication and further progress of AI-driven "emotion" recognition. In response to this terminology problem regarding the term emotion, the Walla Emotion Model (WEM) is put forward as a potentially helpful framework for the next generation of AI. The WEM categorizes affective phenomena into three distinct terms, which are all clearly defined. Those terms are "affection" (i.e., unconscious, neurophysiological responses coding for valence and arousal), "feeling" (conscious, subjective experience), and "emotion" (observable, communicative behavior). Those definitions are meant to allow AI developers to target specific biological and psychological streams, thereby increasing the reliability, validity, and ethical transparency of affective computing.

Keywords: Walla Emotion Model (WEM) · Artificial Intelligence · Emotion Recognition · Neurobiology · Affective Computing · Affection · Feeling · Emotion · Affective information processing · Non-conscious mind · Consciousness · Cognitive pollution

Introduction

We have tasked the most sophisticated pattern-recognition engines in history with identifying a target that remains, even to us, a phantom. To program an AI for "emotion recognition" is like to build a high-resolution map of a territory that refuses to stay still. While a computer requires a binary "yes" or "no" to categorize a state, the incredible number of "emotion" definitions operates more in a spectrum of "perhaps". Before we can ask if an algorithm is accurate in terms of emotion recognition, we must first confront a more haunting question: What exactly is it looking for?

The rise of AI-driven emotion recognition and the related problem

Artificial Intelligence (AI) has moved beyond logical computation into the realm of human affect. Already in 1997, Picard defined affective computing as computing that relates to, arises from, or deliberately influences emotions [1]. In 2010, Calvo and D'Mello published a review on the topic "Affect Detection" [2]. Very critically, they highlighted the problem of multiple views on the definition of the term "emotion" for affective computing and concluded that successful AI systems must be built on specific psychological models that categorize emotions as expressions, physiological embodiments, or social constructs. They emphasize that the most effective systems are multimodal, combining various sensors to resolve the ambiguity of human behavior. Their work acts as a perfect problem statement for this paper by identifying the lack of theoretical consistency in AI, which is just mirroring the lack of consistency in psychology, philosophy, biology and all other fields that are dealing with "emotions" since centuries. The actual problem is that using the term "emotion" as an umbrella term with a varying set of phenomena underneath that differs from one scholar to the other inevitably causes confusion. A programmer does not know what model to follow. As a result of this, different programmers use different models, which obviously creates unwanted chaos and in the worst case causes serious problems.

To date, from automotive safety systems monitoring driver fatigue [3,4] to sentiment analysis in marketing [5], AI-driven "emotion" recognition is a multi-billion-dollar industry. However, the foundational science upon which these systems are built is surprisingly unstable. In the field of psychology alone, there are over 100 competing definitions of "emotion," ranging from Paul Ekman's Basic Emotions [6] to Lisa Feldman Barrett's Theory of Constructed Emotion [7] and many more [8]. This definitional crisis has direct consequences for AI. When a machine is tasked with "recognizing an emotion," it is surprisingly unclear whether it is looking for a physiological process, a social performance, a subjective internal state or something else, all depending on the programmers choice, which emotion model to follow. Without a standardized and clear "emotion" framework that separates the different aspects to it with a clear terminology, AI models suffer from low cross-cultural validity and poor replicability and reliability. Even just talking about functional aspects of algorithms between programmers is difficult.

The data crisis: How definitional ambiguity becomes algorithmic bias

Besides terminological ambiguity, the perhaps most dominant current problem in AI development is largely a result of what is known as the "Ground Truth" problem [9], which initially arose in cartography, where distant data collection needed confirmation through measurements made on the ground. This term has been introduced to AI long ago and became one of its major issues [10]. In supervised machine learning, an algorithm is only as good as the labels provided by human annotators. At the same time, it is well known how self-report is prone to fail, especially when it comes to verbalizing affection-related content [8]. Thus, besides a lack of a clear “emotion” definition, another problem is that most AI datasets rely on self-reporting (asking people how they feel) or human labeling (asking an observer to guess what someone feels based on a picture). Both are inherently subjective and unreliable. Taken together, human trainers cannot agree on what an "emotion" is, and the AI can only ever learn to replicate human bias and linguistic confusion. This is particularly dangerous in high-stakes fields like AI-assisted recruitment or psychiatric diagnosis, where a "misread" emotion can have life-altering consequences (among many more).

Among others, most AI models are trained on datasets like AffectNet [11] or FER2013 [12], which rely on human-coded facial expressions. There is a potential discrepancy between what a person reports they are feeling and what their brain is actually doing, and a facial expression alone might not be a good indication of an underlying affective state (e.g., a fake smile). If an AI is trained on a subjective narrative, it is essentially being trained on a story rather than a biological fact. This leads to affective hallucinations, where the AI predicts an inner state based on a social performance that is not necessarily reflective of a user’s true inner affective state.

Because current AI is based on the conflated terms, unclear terminology and misleading assumptions, it also fails to account for display rules such as cultural norms that dictate how one should look when feeling a certain way. An AI trained in one country might interpret a neutral face in a different culture as sadness or disrespect. By failing to use a model that clearly distinguishes between a behavior and an inner state, developers are inadvertently hard-coding one culture’s emotion behaviors as universal affective truths. In other words, the fundamental problem is that most current AI-driven emotion recognition (ER) systems are built on a categorical misunderstanding. They treat observable behaviors (smiles, scowls) as direct synonyms for internal affective states. If an AI doesn't know what it is actually recognizing, if it confuses emotion with feeling or affection, it creates a disconnect between the data and human reality.

This can indeed lead to very tangible real-world application failures. Just to mention a few, in automotive AI (driver monitoring), a system might interpret a stoic face as alertness, failing to detect the internal actual state of fatigue. In the field of Customer Service & De-escalation, Emotion AI in call centers might flag "rage" based on vocal volume, although it might just be a general high-arousal feeling of excitement. This could cut off a passionate, but loyal customer. An AI-driven mental health app might provide "soothing" feedback based on a user's sad facial expression, while the user’s internal feeling is actually one of pensive reflection or focused problem-solving. Those are just some examples of inappropriate automated interventions.

Additionally, biases and blind spots can occur. The lack of a unified theoretical framework can really introduce significant systematic errors. Most AI is trained on "posed" datasets where actors exaggerate expressions. In the real world, people often experience intense internal affection without any visible emotion. This can create a blind spot for subtle, non- or less expressive individuals, who are then essentially invisible to the AI. Not to forget cultural and contextual blindness. If an AI assumes a 1:1

mapping between a facial movement and a specific feeling, it ignores cultural display rules as has been mentioned above. In some cultures, a smile is a social mask for embarrassment or even anger. An AI without affective depth sees only the mask, but not the actual underlying affective truth. Finally, relying on self-reporting as the ground truth has already been mentioned as potentially problematic. A person's self-report about a felt state or even deep inner affection is only a late-stage cognitive rationalization that might completely misinterpret earlier, non-conscious affection that actually drove their behavior.

The multitude of problems is rather impressive. In order to support further progress in the field of AI-driven emotion recognition as well as to help any involved stakeholders (e.g., healthcare, industries, governments, etc.) it seems useful (if not mandatory) to clarify terminology and respective understanding concerning all terms that are used for the various affective phenomena. A recent review article on emotion recognition and AI also highlighted the problem regarding a lack of a universal definition for the term "emotion" [13]. This is where the Walla Emotion Model can be seen as a potentially helpful framework [8,14]. It provides a very clear terminology that is easy to understand and to communicate and it is rooted in evolutionary neurobiology [15].

The aim of this entire chapter is to provide a theoretical contribution that offers an explained distinction between affective processing, feelings, and emotions with short discussions how this clear terminology improves the validity and interpretability of AI-based emotion recognition? The following section introduces this specific emotion model.

Introducing the Walla Emotion Model: A Neurobiological Solution to Definitional Ambiguity causing Emotion Recognition Problems in AI

The Walla Emotion Model (WEM) [8,14,16] tries to solve the terminological problem by dismantling the word "emotion" (taking away its umbrella character) and by suggesting three distinct, non-overlapping terms that each represent a separate phenomenon related to a single aspect based on neurobiological processing streams [17]. The three distinct terms including short explanations follow below:

1. **Affective processing** (or affection (in contrast to cognition); the unconscious biological truth). It refers to the raw, non-conscious neurophysiological reaction to external or internal stimuli. It occurs in subcortical regions of the brain (limbic system) within milliseconds of a stimulus. It is objective and cannot be faked. For AI, this level of information processing is the "Ground Truth."

2. **Feeling** (The Conscious Subjective Story). Supra-threshold affective processing leads to releases of chemical substances (bodily responses), which are felt in organisms that are capable of consciousness. Feelings are the conscious, mental interpretations of underlying affective processing. They are like perceptions and those are interpreted constructs of the psyche (not one-to-one representations of actual stimuli. The conscious mind often "invents" a story to explain why it is aroused. This makes feelings

interesting for psychology, but potentially dangerous for objective AI training. It is this level, where potential cognitive pollution can influence a felt experience.

3. **Emotion** (The Behavioral Display). According to the WEM [8], an emotion is strictly defined as behavior. This includes facial expressions, body postures, gestures, and vocalizations among many more behaviors that are initially meant to communicate a feeling. Critically, this is only the case for so-called involuntary emotions, but not for voluntary emotions. These can be controlled, faked, or influenced by culture.

In summary, affective processing is understood as the raw data reflecting true evaluative information regarding the valence of a stimulus. A feeling is the subjective and conscious interpretation (perception) of the bodily consequences of affective processing, and finally an emotion is a communicative signal to the outer world.

Interestingly, if one applies this terminology to what AI is currently trained for, the function of "emotion" recognition could keep its label with the only difference that emotion recognition then means nothing but recognizing the behavioral output of affective processing, but certainly not affection itself and also not a feeling. Figure 1 shows a visualization of the WEM.

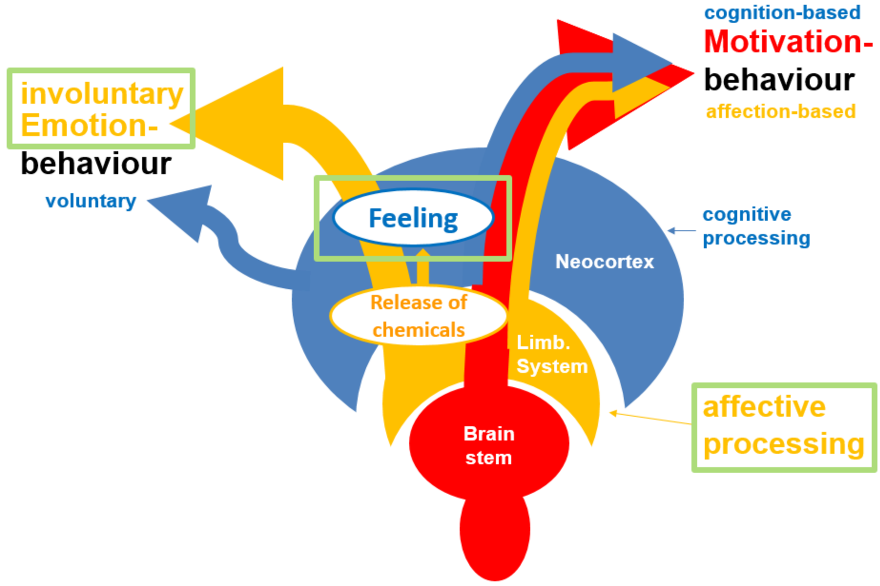


Fig. 1. Visualization of the separate and distinct aspects (or phenomena) of three terms "affective processing", "feeling" and "emotion". Artifacts empowered by Artificial Intelligence (adapted from Walla, 2025 [8].

Advantages of the WEM over traditional Models

Before highlighting potential advantages of the WEM over traditional models, the following paragraph shall provide a short summary across the variety of existing models.

It is not exhaustive, but certainly leaves the impression that an emotion can be many things depending on the model.

The landscape of emotion theory bridges the gap between deep-seated biological drives and complex cognitive interpretations. Charles Darwin [18] initiated this journey by viewing emotional expressions as evolved, functional adaptations, a concept later refined by Carroll Izard [19], who argued for a set of discrete, universal basic emotions. In contrast to purely expressive views, the James-Lange Theory [20, 21] famously proposed that emotions are simply our perception of physiological shifts, while the Schachter-Singer Two-Factor Theory added a crucial layer by stating that we require both physical arousal and a cognitive label to identify what we feel [22]. Richard Lazarus [23] pushed this further with his cognitive appraisal model, asserting that our mental evaluation of a situation's significance is what actually triggers the emotion in the first place. On the neurobiological front, Jaak Panksepp [24] identified seven subcortical "blueprints" for affect (such as SEEKING, FEAR, and RAGE), emphasizing that these are primal, hard-wired systems. This aligns with Jeffrey Gray's BIS-BAS [25] model, which focuses on motivational systems for avoidance (Inhibition) and approach (Activation). Rather than using distinct categories, dimensional theorists like James Russell [26] organize affect along axes of valence and arousal within a circumplex model, a structure often measured by scales like the PANAS [27]. The list could be prolonged, but the few mentioned schools already highlight huge discrepancies and mostly a total lack of a short and clear emotion definition.

This is exactly, where the main advantage of the WEM is best explained. The WEM provides a clear distinction and unambiguous terminology, which is besides supporting emotion research in general also defined as the most important contribution to AI-driven emotion recognition. As a result of applying the WEM's terminology, it becomes clear that "emotion" recognition only means recognizing behavioral consequences of affective processing and certainly not recognizing an inner state of a person, which would be called a feeling according the WEM. Its clear constructs are meant to provide clear targets for AI. Training models on objective physiological data enables the detection of unconscious affective information processing as raw data reflecting true evaluative responses. Using linguistic models to analyze subjective verbal reports, acknowledging their inherent unreliability helps to distinguish between true and false subjective interpretations in a human subject. Similarly, focusing computer vision and audio analysis purely on observable behaviors (facial expressions, tone of voice) might lead to the detection of deceptions as communicative displays.

By adopting the WEM, AI developers can move from unclear and unreliable Emotion Recognition to true Affection Recognition. At least, AI developers can communicate with each other more clearly about affective phenomena they plan to measure and analyze by algorithms. Instead of trying to detect "Anger," an AI can be trained to detect negative affective processing (via physiology) and aggressive emotions (via facial muscles). If the two don't match, the AI identifies a social mask, a level of nuance currently impossible in most models. By separating emotion-behavior (which varies by culture and can be cognitively controlled) from raw affective processing (which simply is the nerve system’s evaluative response to stimuli, independent from culture and subject), AI can be audited for cultural or voluntarily controlled bias.

Implications for NeuroIS Research

In the field of NeuroInformation Systems (NeuroIS), the here highlighted conceptual confusion limits the validity of experimental results. If a NeuroIS study claims to measure user frustration using only facial recognition, it is actually measuring the expression of frustration, not the biological state of frustration. This leads to "noisy" data and inconsistent findings, which only the WEM is able to clearly specify and label. Similarly, by focusing on the conscious feeling (via surveys) rather than the non-conscious biological affection, NeuroIS research may miss the true drivers of technology adoption or user friction. In the end, it does seem potentially dangerous trying to perform mind-reading, but failing to do so, while adapting a system the wrong way. The assumption that because an AI can detect an emotion (according to the WEM only the behavioral expression), it has access to a person's feeling (according to the WEM the only affective phenomenon reflecting a subjective experience) could lead to invasive corporate or governmental policies based on "reading" people who are actually just having a neutral biological reaction.

Conclusion: Towards an Ethical and Robust AI

The current problem for AI-driven “emotion” recognition is not a lack of data, but a lack of conceptual clarity. As long as AI treats "emotion" as a single, messy category, it will remain an unreliable tool. The WEM allows us to map AI capabilities to the brain's evolutionary palimpsest. By training machines to distinguish between the biological truth of affection, the subjective narrative of a feeling, and the behavioral display of an emotion, we can build systems that are not only more accurate, but more ethically transparent. In an era where "Brain Rot", the word of the year 2024 selected by experts from the Oxford University Press [28], most likely caused by "Post-Truth" (word of the year 2016) [29], threatens to degrade human cognition, an AI that understands the difference between a raw biological reaction and a performed emotion is essential for the future of human-machine interaction. This paper concludes that the so-called "Affective Turn" [30] must be met with a "Neurobiological Turn" in AI architecture. By implementing the WEM, we could move from a world of simulated empathy (AI guessing at social displays) to biological synchrony (AI understanding the actual affective state of a human). This transition is the only way to prevent AI from accelerating the cognitive decline of society and instead use it as a tool to support human cognitive health.

While this is all pure speculation, the proposed WEM is primarily meant to serve as a vital tool for the AI community, offering a neurobiological clean-up of the term emotion that aligns machine learning targets with actual human brain architecture.

References

1. Picard, R. W. (1997). Affective computing. The MIT Press.
2. Calvo, R., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing, 1, 18-37.
3. Visconti, P., Rausa, G., Del-Valle-Soto, C., Velázquez, R., Cafagna, D., & De Fazio, R. (2025). Innovative Driver Monitoring Systems and On-Board-Vehicle Devices in a Smart-Road Scenario Based on the Internet of Vehicle Paradigm: A Literature and Commercial Solutions Overview. Sensors, 25(2), 562.
4. Yang, E., & Yi, O. (2024). Enhancing Road Safety: Deep Learning-Based Intelligent Driver Drowsiness Detection for Advanced Driver-Assistance Systems. Electronics, 13(4), 708.
5. Maghsoudi M, Bakhtiari M, Bakhtiari H (2026) A hybrid framework for notebook market analysis: Integrating social media sentiment mining with expert knowledge for feature prioritization. PLOS ONE 21(2): e0342067.
6. Ekman, P. An argument for basic emotions. Cogn. Emot. 1992, 6, 169–200.
7. Barrett, L.F. The theory of constructed emotion: An active inference account of interoception and categorization. Soc. Cogn. Affect. Neurosci. 2017, 12, 1–23.
8. Walla, P., Wolman, A., & Northoff, G. (2025). A Call for Conceptual Clarity: "Emotion" as an Umbrella Term Did Not Work—Let's Narrow It Down. Brain Sciences, 15(9), 929.
9. Garrity, G.M. (2009). Ground truth. Stand Genomic Sci; 1(2): 91-2.
10. Girard-Chanudet, C. (2025). Ground-truth is law: The invisible conceptual work behind AI. Big Data & Society, 12(2).
11. Mollahosseini, A., Hasani, B., and Mahoor, M.H. (2019). "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild" in IEEE Transactions on Affective Computing, 10(1): 18-31.
12. Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., Ionescu, R., ... Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. arXiv.
13. Khare, S.K., Blanes-Vidal, V., Nadimi, E.S., and Acharya, U.R. (2024). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. Information Fusion, 102: 102019.
14. Walla, P. (2018). Affective Processing Guides Behavior and Emotions Communicate Feelings: Towards a Guideline for the NeuroIS Community. In Information Systems and Neuroscience; Davis, F., Riedl, R., vom Brocke, J., Léger, P.M., Randolph, A., Eds.; Lecture Notes in Information Systems and Organisation; Springer: Cham, Switzerland; Volume 25.
15. Walla, P., and Panksepp, J. (2013). Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions. Novel Frontiers of Advanced Neuroimaging, Kostas N. Fountas (Ed.), ISBN: 978-953-51-0923-5, InTech.
16. Walla, P. (2026). Utilizing the Walla Emotion Model to Standardize Terminological Clarity for AI-Driven "Emotion" Recognition. Brain Sciences, 16(3), 260.
17. Walla, P. (2026). The Olfactory Origins of Affective Processing: A Neurobiological Synthesis Through the Walla Emotion Model. Life, 16(1), 86.
18. Darwin, C. (1872). The Expression of the Emotions in Man and Animals; John Murray: London, UK.
19. Izard, C.E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. Annu. Rev. Psychol., 60, 1–25.
20. James, W. (1884). What is an emotion? Mind, 9, 188–205.

21. Lange, C.G. (1885). The mechanism of the emotion. (B. Rand, Trans.). In The Classical Psychologists; Rand, B., Ed.; University Press of the Pacific: Forest Grove, OR, USA,; pp. 672–684.
22. Schachter, S.; Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. Psychol. Rev., 69, 370–399.
23. Lazarus, R.S. (1991). Emotion and Adaptation; Oxford University Press: Oxford, UK.
24. Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? Neurosci. Biobehav. Rev., 35, 1791–1804.
25. Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology, 67(2), 319–333.
26. Russell, J.A. (1980). A circumplex model of affect. J. Personal. Soc. Psychol., 39, 1161–1178.
27. 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.
28. Oxford University Press. (2024). Oxford Word of the Year 2024: Brain rot. Oxford Languages.
29. Oxford University Press. (2016). Word of the Year 2016: Post-truth. Oxford Languages.
30. Wahl-Jorgensen, K. (2020). An Emotional Turn in Journalism Studies? Digital Journalism, 8(2), 175–194.

The Slow Drift of Alpha/Mu Rhythms During Sustained Motor Engagement

Kyriaki Kostoglou[1*], Gernot R. Müller-Putz[1,2]

[1] Institute of Neural Engineering, Graz University of Technology, Graz, Austria
[2] BioTechMed, Graz, Austria
kkostoglou@tugraz.at

Abstract. Brain oscillations undergo slow adaptations during prolonged task engagement that are not captured by conventional power-based electroencephalography (EEG) features. Using a large-scale, publicly available EEG dataset recorded during sustained motor engagement, we analyzed session-long alpha/mu dynamics with a recursive tracking framework. Alpha/mu frequency showed acceleration over sensorimotor regions and slowing in surrounding non-sensorimotor areas, while magnitude increased broadly with a sensorimotor emphasis. Clustering of frequency–trend topographies revealed distinct adaptation profiles, highlighting oscillatory frequency as a complementary dimension of neural adaptation for adaptive EEG-based systems.

Keywords: Alpha slowing, Mu acceleration, frequency tracking, autoregressive, movement execution, motor imagery

Introduction

Electroencephalography (EEG) has become a central tool for examining how brain activity evolves during prolonged interaction, learning, and task performance. In research areas spanning brain–computer interfaces (BCIs), neuroergonomics, and neuro-information systems (NeuroIS), brain signals are typically summarized using measures of signal magnitude or power within predefined frequency bands. Such measures have been instrumental in characterizing engagement, stress, workload, and fatigue [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], and they support a wide range of neuroadaptive systems. However, power-based descriptions primarily capture the prominence of neural rhythms and provide only a partial account of how brain dynamics evolve over time. In particular, oscillatory brain activity is not spectrally static: the center frequency of canonical rhythms can shift over time in response to sustained cognitive and motor demands, learning, and changes in internal state. For example, the peak frequency of alpha activity has been shown to vary with arousal, vigilance, and cognitive load, often decreasing during mental fatigue and increasing during heightened alertness or task engagement [12, 13, 14, 15, 16, 17, 18]. Similarly, sensorimotor mu rhythms can exhibit frequency changes during motor learning or repeated motor imagery, reflecting evolving synchronization within sensorimotor networks [19, 20]. From a neurophysiological perspective, such frequency shifts reflect changes in the coordination and interaction of neuronal populations, which shape the timing and efficiency of information exchange across brain networks [21].

Recently, we introduced and validated a state-space oscillator tracking framework capable of continuously estimating instantaneous EEG frequency and magnitude, enabling the characterization of time-varying oscillatory dynamics beyond conventional spectral measures [22]. This approach builds upon and extends our previous work on pole-tracking based EEG analysis [23, 24, 25], and enables the characterization of gradual, session-long changes in oscillatory dynamics, including frequency trends that capture long-term neural adaptation beyond instantaneous fluctuations. In [22], we applied this framework across multiple datasets to explore changes in alpha/mu rhythms, which play a central role in motor-related EEG activity [26], during motor-related BCI tasks. Although alpha and mu rhythms share the same frequency range, they differ in their spatial distribution and functional characteristics, with mu activity localized over sensorimotor regions and modulated by movement-related processes, whereas alpha activity is more prominent in parieto-occipital regions and reflects broader cognitive and vigilance-related states. As such, their separation based on frequency alone is challenging and requires consideration of spatial and functional information [27]. In [22], we observed consistent but opposing frequency trends during motor-related BCI tasks, with sensorimotor mu acceleration and concurrent alpha slowing in non-sensorimotor regions. These opposing dynamics were interpreted as reflecting distinct but co-occurring neural processes: mu acceleration over sensorimotor regions linked to increased neural efficiency, engagement, or training-related plasticity, and alpha slowing in non-sensorimotor areas associated with fatigue, reduced vigilance, or functional downregulation of task-irrelevant networks.

In the present study, we go one step further by focusing on the largest dataset among those previously examined, enabling a more detailed and robust characterization of oscillatory dynamics during sustained task engagement. Beyond group-level patterns, we further explore inter-individual variability by clustering participants according to their spatial patterns of these dynamics. This allows us to identify distinct modes of cortical adaptation, moving beyond a single averaged representation toward a more comprehensive and individualized account of neural responses.

Materials and Methods

Dataset and Preprocessing

We analyzed EEG data from a large publicly available dataset of 109 healthy participants [28] (https://doi.org/10.13026/C28G6P), including motor execution and imagery tasks of unilateral and bilateral hand and foot movements. Each recording session lasted approximately 25 minutes. To ensure consistency, 9 shorter recordings were excluded, yielding a final sample of 100 participants. EEG was recorded from 64 channels (10–20 system) at 160 Hz. Data were high-pass filtered at 0.5 Hz (4th order Butterworth), and notch filtered at 60 Hz. Independent component analysis (extended Infomax algorithm [29], EEGLAB [30]) was applied to remove eye and muscle artifacts. The signals were then band-pass filtered in the alpha/mu range (8–12 Hz).

Alpha/Mu Tracking

The tracking approach [22] starts by modeling the band-pass–filtered EEG signal $\boldsymbol{y}$ as a second-order autoregressive (AR(2)) oscillator [31],

$$y(n) = a_1(n)y(n-1) + a_2(n)y(n-2) + v(n) \tag{1}$$

where the current sample is expressed as a linear combination of the two previous samples plus noise ($\boldsymbol{v}$). For oscillatory signals, an AR(2) model acts as a discrete-time resonator characterized with complex-conjugate poles, expressed by a radius $r(n)$ and an angle $\omega(n)$, which encode damping and instantaneous angular frequency, respectively, via the following reparameterization,

$$a_1(n) = 2rcos(\omega_n T_s),\ \ a_2(n) = -r(n)^2 \tag{2}$$

This reparameterization yields a nonlinear state-space resonator with physiologically interpretable states. The latent state includes internal resonator components that generate the oscillatory signal and slowly varying parameters governing instantaneous frequency and damping. An Extended Kalman Filter [22, 23, 32] is used to track gradual temporal changes in these states under noise and nonlinearity. To ensure reliable performance, the EKF hyperparameters, including noise covariances, initial conditions, and state uncertainty, are tuned using a genetic algorithm [22, 23, 33, 34, 35]. The optimization is performed on a short calibration segment by minimizing a prediction error criterion aggregated across channels, after which the selected hyperparameters are fixed for the full recording. Instantaneous frequency is obtained from the tracked pole angle, while instantaneous magnitude is derived from the resonator state components. Implementation of the algorithm is available in a public GitHub repository [36].

Session-long trends in alpha/mu frequency and magnitude were assessed by correlating (Pearson's correlation) the tracked trajectories with elapsed recording time, allowing the identification of gradual changes in oscillatory dynamics. This analysis was performed independently for each EEG channel, and the resulting correlation values were then averaged across participants to obtain group-level topographical maps of alpha/mu frequency and magnitude trends.

Clustering

To separate individuals into groups based on the spatial organization of alpha/mu frequency acceleration and deceleration across the scalp, we focused on subject-level topographical maps of alpha/mu frequency trends, as these directly reflect changes in oscillatory timing and were of primary interest in this study. Unsupervised clustering was performed using *k*-medoids [37], with similarity between subjects quantified using either correlation distance, which emphasizes relative spatial patterns of frequency change, or squared Euclidean distance, which additionally accounts for differences in the overall strength of frequency trends. The optimal number of clusters was evaluated using both the gap statistic [38] and the silhouette criterion [39].

Results

Figure 1a shows group-level scalp topographies of correlations between elapsed recording time and alpha/mu frequency (left) and magnitude (right). Alpha/mu frequency exhibited a clear spatial structure, with positive correlations over central and centroparietal regions and negative correlations over posterior and peripheral areas, indicating regional frequency acceleration and slowing across the session. In contrast, alpha/mu magnitude increased broadly across the scalp, with stronger positive correlations over central and centroparietal regions.

Figure 1b presents clustering results for subject-level alpha/mu frequency–trend topographies using correlation distance (left) and squared Euclidean distance (right). In all cases, the optimal number of clusters was two (Cluster A and Cluster B). Correlation-based clustering emphasized a top–bottom opposition, separating clusters by frontal–central (Cluster A; N=48) versus centroparietal-posterior (Cluster B; N=52) frequency changes, whereas squared Euclidean clustering highlighted a more lateralized, contralateral organization of frequency adaptation across hemispheres (Cluster A; N = 45, Cluster B; N = 55).

Discussion

We employed a state-space oscillator tracking framework to estimate instantaneous frequency and magnitude of alpha/mu rhythms, capturing dynamics beyond conventional spectral measures. While the framework is capable of capturing both short- and long-term changes in oscillatory activity, here we focused specifically on long-term adaptations by relating the tracked features to elapsed time, thereby quantifying slow, session-level trends. This makes the approach particularly well suited for sustained, nonstationary scenarios (such as fatigue, workload, or learning) where gradual neural changes are of primary interest. At the same time, by providing instantaneous estimates, the method is naturally suited for real-time monitoring and adaptive systems, offering a clear advantage over Hilbert-based or time–frequency approaches that rely on windowing or batch processing and introduce latency.

Using this approach, we observed systematic, session-long changes in alpha/mu oscillatory dynamics during sustained motor tasks, with distinct spatial patterns for frequency and magnitude. At the group level, alpha/mu frequency increased over central/centroparietal sensorimotor regions, while concurrent decreases were observed in surrounding non-sensorimotor areas. In contrast, alpha/mu magnitude increased broadly across the scalp, with a stronger emphasis over centroparietal regions. The acceleration of alpha/mu frequency over sensorimotor regions may have reflected increasingly efficient or tightly synchronized motor-related processing as the session

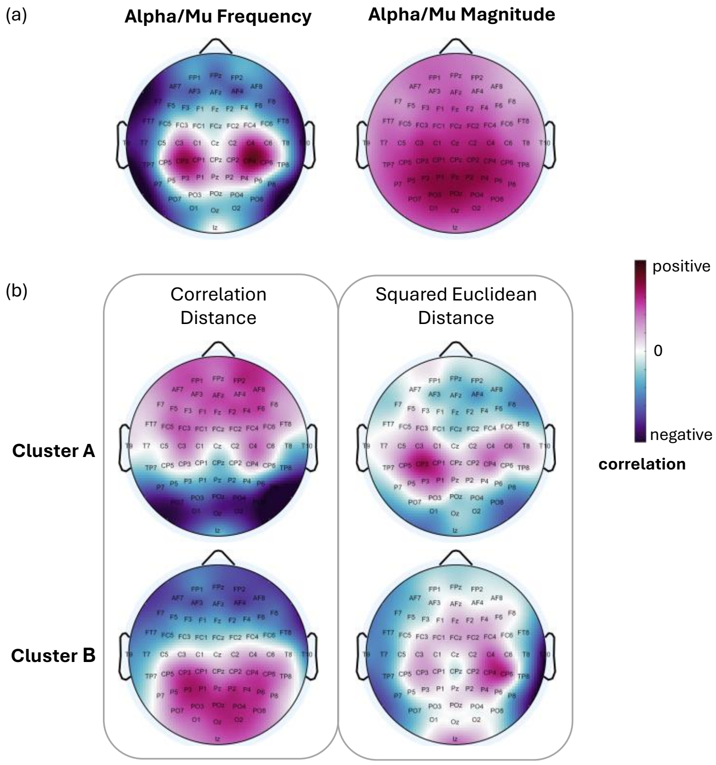


Fig. 18. (a) Grand-average scalp topographies across participants showing the correlation between elapsed recording time and alpha/mu oscillatory frequency (left) and magnitude (right). Positive correlations indicate gradual increases over the session, while negative correlations indicate decreases. (b) Cluster-averaged scalp maps of alpha/mu frequency trends obtained using *k*-medoids clustering with correlation distance (left) and squared Euclidean distance (right). Two clusters were identified: Cluster A (top) and B (bottom).

progressed [19, 22, 40]. Pfurtscheller [41] demonstrated that with training, upper mu (10–12 Hz) event-related desynchronization (ERD), a task-related attenuation of oscillatory power reflecting sensorimotor cortical activation, becomes more spatially localized, suggesting more functionally specific neural recruitment. This spatial refinement may facilitate faster and more coherent oscillatory activity in task-relevant motor areas, contributing to the observed mu frequency acceleration. In this context, higher frequency may reflect increased excitability and stronger coordination within motor networks. On the other hand, frequency slowing in surrounding non-sensorimotor areas may have been associated with global state changes such as fatigue and disengagement [17, 42, 43] or with functional downregulation of task-irrelevant networks during sustained motor engagement [44]. The widespread increase in magnitude, particularly over sensorimotor areas, likely reflected stabilization or strengthening of oscillatory activity rather than reduced engagement [22]. These interpretations align with prior work but should be viewed cautiously, as no direct behavioral or physiological validation was provided here. Nevertheless, the observed effects are not arbitrary. In [22], we demonstrated that frequency changes were coordinated across sensorimotor and non-sensorimotor regions, indicating coherent network-level interactions rather than independent fluctuations.

Beyond group-level trends, a novel contribution of the present work was the identification of distinct modes of cortical adaptation during sustained motor engagement through clustering of subject-level alpha/mu frequency–trend topographies. Correlation-based clustering, which emphasizes relative spatial organization, identified two dominant patterns characterized by an anterior–posterior opposition. In one mode, frequency increased over frontal and central sensorimotor regions and decreased over posterior non-sensorimotor areas. This pattern may be indicative of a more top–down–dominated form of adaptation, where task-relevant motor and premotor networks operate at relatively faster oscillatory timescales (potentially reflecting increased temporal precision, excitability, or reliance on internally driven motor representations), while posterior regions exhibit relative downregulation. In the other mode, frequency acceleration occurred over posterior regions alongside slowing over frontal–central areas. This may suggest a bottom–up–weighted or sensory-driven adaptation strategy, in which posterior sensory or visuospatial regions maintain higher temporal resolution to support sensory monitoring or imagery while frontal and sensorimotor regions operate at slower timescales. These opposing spatial patterns align with previously described large-scale alpha-frequency organization axes and anterior–posterior gradients reported in resting-state and task- related EEG [45], suggesting that individuals may differ in how such network-level frequency coordination mechanisms are expressed and accumulated over prolonged engagement. The lateralized patterns emphasized by squared Euclidean clustering pointed to additional hemispheric asymmetries in frequency adaptation. Such contralateral organization may have reflected differences in motor strategy, handedness, or imagery vividness, with stronger frequency adaptation occurring in the hemisphere most engaged by the task. These asymmetries may also reflect inter-individual variability in sensorimotor lateralization, which is known to vary even among healthy participants performing identical tasks. Although these interpretations are intriguing, the underlying functional mechanisms are not directly investigated in the present study.

Together, these findings indicate that alpha/mu frequency adaptation is not a unitary process but reflects multiple, coexisting neural mechanisms that vary across individuals. For NeuroIS, this means that sustained human–technology interaction leads to different neural adaptation patterns across individuals, supporting more personalized system design instead of one-size-fits-all approaches. Moreover, while most work has focused on oscillatory power, our findings highlight frequency dynamics as a complementary marker of long-term adaptation. In this context, frequency-based features are particularly suited for tracking gradual state changes, such as fatigue, workload, prolonged decision-making, sustained attention and may also provide insight into users' affective states during human–technology interaction. For example, in air-traffic control [46] or driving [47], frequency slowing may signal declining vigilance before behavioral errors occur, enabling early intervention. Similarly, in digital or immersive environments (e.g., virtual agents or avatars [48]), frequency dynamics could track users' affective states during interaction [49] and guide adaptive responses. In digital work or human–AI interaction, they could inform interface adaptation based on cognitive load. Unlike power, frequency captures temporal coordination and may provide earlier or more sensitive indicators of state transitions and performance decline. In training contexts, such as BCI or skill acquisition, it could potentially help detect shifts between engagement and fatigue and support adaptive adjustments in task difficulty or feedback to optimize learning. Incorporating frequency tracking could also enable dynamic adaptation of feature spaces, for example by adjusting frequency bands in real time to match individual or time-varying peak frequencies. This may be particularly beneficial in applications such as adaptive BCI control, where fixed-band assumptions may fail under prolonged use. Moreover, frequency dynamics may help disentangle overlapping neural processes, as similar power changes can arise from different underlying mechanisms, whereas shifts in oscillatory timing may reflect more specific changes in neural coordination or network state.

Although the present study focused on sustained motor-related engagement, the proposed frequency-tracking framework is not task-specific. Frequency shifts in the alpha band have also been observed during prolonged cognitive workload, attentional control, mental fatigue, and sustained decision-making [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], paradigms that are central to NeuroIS research. This indicates that the approach is relevant across a wide range of sustained interaction scenarios, as outlined earlier. Importantly, the underlying tracking approach is compatible with online implementation, providing a pathway toward real-time monitoring, which remains challenging for many existing EEG-based methods.

References

1. Wriessnegger, S.C., Raggam, P., Kostoglou, K., Müller-Putz, G.R.: Mental State Detection Using Riemannian Geometry on Electroencephalogram Brain Signals. Front. Hum. Neurosci. 15, (2021). https://doi.org/10.3389/fnhum.2021.746081.
2. Riedl, R., Kostoglou, K., Wriessnegger, S.C., Müller-Putz, G.R.: Videoconference fatigue from a neurophysiological perspective: experimental evidence based on electroencephalography (EEG) and electrocardiography (ECG). Sci. Rep. 13, 18371 (2023). https://doi.org/10.1038/s41598-023-45374-y.
3. Wriessnegger, S.C., Lorenzer, L., Kostoglou, K.: Neurocardiac signatures of acute mental stress: a sex-comparative study. Front. Neurosci. 19, 1633295 (2025). https://doi.org/10.3389/fnins.2025.1633295.
4. Wriessnegger, S.C., Leitner, M., Kostoglou, K.: The brain under pressure: Exploring neurophysiological responses to cognitive stress. Brain Cogn. 182, 106239 (2024). https://doi.org/10.1016/j.bandc.2024.106239.
5. Egger, J., Kostoglou, K., Müller-Putz, G.R.: Chrono-EEG dynamics influencing hand gesture decoding: a 10-hour study. Sci. Rep. 14, 20247 (2024). https://doi.org/10.1038/s41598-024-70609-x.

6. Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., Tsiknakis, M.: Review on Psychological Stress Detection Using Biosignals. IEEE Trans. Affect. Comput. 13, 440–460 (2022). https://doi.org/10.1109/TAFFC.2019.2927337.
7. Kakkos, I., Dimitrakopoulos, G.N., Sun, Y., Yuan, J., Matsopoulos, G.K., Bezerianos, A., Sun, Y.: EEG fingerprints of task-independent mental workload discrimination. IEEE J. Biomed. Health Inform. 25, 3824–3833 (2021). https://doi.org/10.1109/JBHI.2021.3085131.
8. Jap, B.T., Lal, S., Fischer, P., Bekiaris, E.: Using EEG spectral components to assess algorithms for detecting fatigue. Expert Syst. Appl. 36, 2352–2359 (2009). https://doi.org/10.1016/j.eswa.2007.12.043.
9. Apicella, A., Arpaia, P., Frosolone, M., Improta, G., Moccaldi, N., Pollastro, A.: EEG-based measurement system for monitoring student engagement in learning 4.0. Sci. Rep. 12, 5857 (2022). https://doi.org/10.1038/s41598-022-09578-y.
10. Käthner, I., Wriessnegger, S.C., Müller-Putz, G.R., Kübler, A., Halder, S.: Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain--computer interface. Biol. Psychol. 102, 118–129 (2014). https://doi.org/10.1016/j.biopsycho.2014.07.014.
11. Dimitriadis, S.I., Sun, Y.U., Kwok, K., Laskaris, N.A., Thakor, N., Bezerianos, A.: Cognitive workload assessment based on the tensorial treatment of EEG estimates of cross-frequency phase interactions. Ann. Biomed. Eng. 43, 977–989 (2015). https://doi.org/10.1007/s10439-014-1143-0.
12. Osaka, M.: Peak alpha frequency of EEG during a mental task: Task difficulty and hemispheric differences. Psychophysiology. 21, 101–105 (1984). https://doi.org/10.1111/j.1469-8986.1984.tb02325.x.
13. Buccellato, A., Zang, D., Zilio, F., Gomez-Pilar, J., Wang, Z., Qi, Z., Zheng, R., Xu, Z., Wu, X., Bisiacchi, P., others: Disrupted relationship between intrinsic neural timescales and alpha peak frequency during unconscious states–a high-density EEG study. Neuroimage. 265, 119802 (2023). https://doi.org/10.1016/j.neuroimage.2022.119802.
14. Klimesch, W., Sauseng, P., Hanslmayr, S.: EEG alpha oscillations: the inhibition–timing hypothesis. Brain Res. Rev. 53, 63–88 (2007). https://doi.org/10.1016/j.brainresrev.2006.06.003.
15. Klimesch, W.: EEG-alpha rhythms and memory processes. International Journal of psychophysiology. 26, 319–340 (1997). https://doi.org/10.1016/s0167-8760(97)00773-3.
16. Angelakis, E., Lubar, J.F., Stathopoulou, S., Kounios, J.: Peak alpha frequency: an electroencephalographic measure of cognitive preparedness. Clinical Neurophysiology. 115, 887–897 (2004). https://doi.org/10.1016/j.clinph.2003.11.034.
17. Benwell, C.S.Y., London, R.E., Tagliabue, C.F., Veniero, D., Gross, J., Keitel, C., Thut, G.: Frequency and power of human alpha oscillations drift systematically with time-on-task. Neuroimage. 192, 101–114 (2019). https://doi.org/10.1016/j.neuroimage.2019.02.067.
18. Jacques, C., Verdonk, C., Cardoso, E., Ferreira, A., Vanneau, T., Beauchamps, V., Dondaine, T., Léger, D., Le Van Quyen, M., Gomez-Merino, D., others: Successful Closed-loop Neurofeedback Alpha Frequency Modulation Enhances the Temporal Dynamics of Attention. Neuroimage. 121912 (2026). https://doi.org/10.1016/j.neuroimage.2026.121912.
19. Di Nota, P.M., Chartrand, J.M., Levkov, G.R., Montefusco-Siegmund, R., DeSouza, J.F.X.: Experience-dependent modulation of alpha and beta during action observation and motor imagery. BMC Neurosci. 18, 28 (2017). https://doi.org/10.1186/s12868-017-0349-0.

20. Hülsdünker, T., Mierau, A., Strüder, H.K.: Higher balance task demands are associated with an increase in individual alpha peak frequency. Front. Hum. Neurosci. 9, 695 (2016). https://doi.org/10.3389/fnhum.2015.00695.
21. Buzsaki, G., Draguhn, A.: Neuronal oscillations in cortical networks. Science (1979). 304, 1926–1929 (2004). https://doi.org/10.1126/science.1099745.
22. Kostoglou, K., Müller-Putz, G.R.: Opposing cortical forces: Alpha slowing and sensorimotor mu acceleration during motor-related BCI training. PLoS Comput. Biol. 22, e1014112 (2026). https://doi.org/10.1371/journal.pcbi.1014112.
23. Kostoglou, K., Müller-Putz, G.R.: Motor-Related EEG Analysis Using a Pole Tracking Approach. IEEE Transactions on Neural Systems and Rehabilitation Engineering. (2024). https://doi.org/10.1109/TNSRE.2024.3483294.
24. Kostoglou, K., Lunglmayr, M.: Root tracking using time-varying autoregressive moving average models and sigma-point Kalman filters. EURASIP J. Adv. Signal Process. 2020, 6 (2020). https://doi.org/10.1186/s13634-020-00666-7.
25. Müller-Putz, G., Crell, M., Egger, J., Suwandjieff, P., Kostoglou, K.: Towards Implantable Brain-Computer Interface for Communication in Locked-In Syndrome patients: An introduction to INTRECOM. In: Current Directions in Biomedical Engineering. pp. 1–4 (2023). https://doi.org/https://doi.org/10.1515/cdbme-2023-1201.
26. Pfurtscheller, G., Da Silva, F.H.L.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clinical neurophysiology. 110, 1842–1857 (1999). https://doi.org/10.1016/s1388-2457(99)00141-8.
27. Garakh, Z., Novototsky-Vlasov, V., Larionova, E., Zaytseva, Y.: Mu rhythm separation from the mix with alpha rhythm: Principal component analyses and factor topography. J. Neurosci. Methods. 346, 108892 (2020). https://doi.org/10.1016/j.jneumeth.2020.108892.
28. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans. Biomed. Eng. 51, 1034–1043 (2004). https://doi.org/10.1109/TBME.2004.827072.
29. Lee, T.-W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput. 11, 417–441 (1999). https://doi.org/10.1162/089976699300016719.
30. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods. 134, 9–21 (2004). https://doi.org/10.1016/j.jneumeth.2003.10.009.
31. Hamilton, J.D.: Time series analysis. Princeton university press Princeton (1994).
32. Ljung, L.: Asymptotic behaviour of extended Kalman filter as a parameter estimator for linear systems. IEEE Trans. on Auto. Control. AC-24, 36–50 (1979). https://doi.org/10.1109/TAC.1979.1101943.
33. Kostoglou, K., Schondorf, R., Mitsis, G.D.: Modeling of multiple-input, time-varying systems with recursively estimated basis expansions. Signal Processing. 155, 287–300 (2019). https://doi.org/10.1016/j.sigpro.2018.09.040.
34. Kostoglou, K., Müller-Putz, G.R., Graz, B.: Optimizing time-varying autoregressive models for BCI applications. In: 9th Graz Brain-Computer Interface Conference: Join Forces-Increase Performance: BCI 2024. pp. 17–22 (2024). https://doi.org/10.3217/978-3-99161-014-4-004.

35. Kostoglou, K., Robertson, A.D., MacIntosh, B., Mitsis, G.D.: A novel framework for estimating time-varying multivariate autoregressive models and application to cardiovascular responses to acute exercise. IEEE Trans. Biomed. Eng. (2019). https://doi.org/10.1109/TBME.2019.2903012.
36. K. Kostoglou: EKF Oscillator Tracking, https://github.com/graz-bci/EKF-Oscillator-Tracking.
37. Park, H.-S., Jun, C.-H.: A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl. 36, 3336–3341 (2009). https://doi.org/10.1016/j.eswa.2008.01.039.
38. Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Series B Stat. Methodol. 63, 411–423 (2001). https://doi.org/10.1111/1467-9868.00293.
39. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. (1987). https://doi.org/10.1016/0377-0427(87)90125-7.
40. Li, B.-Z., Nan, W., Pun, S.H., Vai, M.I., Rosa, A., Wan, F.: Modulating individual alpha frequency through short-term neurofeedback for cognitive enhancement in healthy young adults. Brain Sci. 13, 926 (2023). https://doi.org/10.3390/brainsci13060926.
41. Pfurtscheller, G.: Spatiotemporal analysis of alpha frequency components with the ERD technique. Brain Topogr. 2, 3–8 (1989). https://doi.org/10.1007/BF01128838.
42. Rubio-Morales, A., Diaz-Garcia, J., Berchicci, M., Morenas-Martin, J., Del Campo, V.L., Garcia-Calvo, T.: The Mental Fatigue Induced by Physical, Cognitive and Combined Effort in Amateur Soccer Players: A Comparative Study Using EEG. J. Funct. Morphol. Kinesiol. 10, 373 (2025). https://doi.org/10.3390/jfmk10040373.
43. Arnau, S., Brümmer, T., Liegel, N., Wascher, E.: Inverse effects of time-on-task in task-related and task-unrelated theta activity. Psychophysiology. 58, e13805 (2021). https://doi.org/10.1111/psyp.13805.
44. Jensen, O., Mazaheri, A.: Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4, 186 (2010). https://doi.org/10.3389/fnhum.2010.00186.
45. Smith, M.K., Grabowecky, M., Suzuki, S.: Dynamic formation of a posterior-to-anterior peak-alpha-frequency gradient driven by two distinct processes. eNeuro. 11, (2024). https://doi.org/10.1523/ENEURO.0273-24.2024.
46. Dasari, D., Shou, G., Ding, L.: ICA-derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task. Front. Neurosci. 11, 297 (2017). https://doi.org/10.3389/fnins.2017.00297.
47. Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., Caltagirone, C., Kong, W., Wei, D., Zhou, Z., others: Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. In: 2012 annual international conference of the IEEE engineering in medicine and biology society. pp. 6442–6445 (2012).
48. Sylaiou, S., Fidas, C.: Virtual humans in museums and cultural heritage sites. Applied Sciences. 12, 9913 (2022).
49. Mühl, C., Allison, B., Nijholt, A., Chanel, G.: A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. Brain-Computer Interfaces. 1, 66–84 (2014).

Neurophysiological Signatures of LLM Hallucinations and Their Emotional Impact on Student Performance: A Research-in-Progress Study

Pramit Khatua[1*], Prof. Himanshu Gupta[2]
[1]Research Scholar, Department of Management Studies and Industrial Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad
(22dr0165@iitism.ac.in)
[2]Assistant Professor, Department of Management Studies and Industrial Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad
(himanshuguptadoms@gmail.com)

Abstract. Large language models are increasingly used by students for learning and task completion. Yet LLM hallucinations, fluent but incorrect or unsupported outputs, can provoke negative emotions and may impair student performance through disrupted attention, reduced confidence, and inefficient verification behaviors. This research-in-progress study investigates the emotional impact of hallucinations on student task performance using simultaneous EEG and Tobii eye tracking to capture in-situ affective and attentional responses beyond self-reports. We collected initial multimodal recordings from 32 students interacting with an LLM during an academic task. We extract EEG spectral features and eye metrics aligned to hallucination episodes. We then compare Decision Tree, Random Forest, and XGBoost models for classifying emotional response states and predicting downstream performance outcomes. The study contributes to emotions in human-AI interaction and offers design implications for safer, emotion-aware LLM support in education.

Keywords: LLM hallucination · emotion · student performance EEG · eye tracking · machine learning

Introduction

The rapid inclusion of Large Language Models into the workflow of education (as a tool for generating ideas, explaining concepts, summarizing information, and providing support for writing) has the potential to increase the efficiency and effectiveness of many educational tasks. However, the output of these models can include hallucinations, outputs that contain information that appears to be true and certain but actually does not exist, is fabricated, or is based on insufficient data. Hallucinations have the potential to impact the educational process negatively in three ways. Students who use hallucinations will potentially consume false or inaccurate information and therefore miss out on learning important information [1]. Students will potentially expend a great deal of time and effort attempting to verify or correct the false or inaccurate information produced by the model. Students' affective response to hallucinations can cause students to become less persistent and perform less well than they would otherwise [3]. The first challenge facing researchers is how to measure the affective response of students to hallucinations. Students' affective responses to hallucinations typically occur very quickly and are often experienced unconsciously, which makes it difficult to measure their affective response through self-report alone. Self-report is supplemented by the use of psychophysiological measures of affective response that provide a means to assess students' affective responses, i.e., when the affective response occurs [4]. At the same time, the methodology indicates that there are several factors to take into account when selecting a physiological instrument for assessing students' affective responses to hallucinations. These factors include reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness [7]. We have included a brief affect calibration block using a Positive-Neutral-Negative (PNN) framework before the LLM task to enhance the construct validity of the study by allowing us to determine within participant reference patterns for affective arousal changes which will allow us to use self-report data as an additional method of validation when participants experience hallucinations.

This paper presents a work-in-progress study that uses concurrent EEG and Tobii eye-tracking to examine how hallucinations are associated with changes in affective and attentional signatures and whether the signatures of affective and attentional responses to hallucinations are related to the performance of students. Both EEG and eye-tracking are two of the most frequently used tools in NeuroIS research [2]. Each provides high levels of temporal resolution and supports the conduct of ecological interaction studies.

Literature Support

Multimodal measurement of affect and attention

The temporal resolution of EEG allows for millisecond level detail in which to analyze data via spectral techniques (frequency bands) that specifically measure how a brain responds to an event as it occurs [10]. Eye tracking captures where attention is allocated on-screen and is well-suited for rapidly unfolding visual inspection processes. However, eye tracking is very well suited for measuring where a person's visual inspection is focused during rapid on-screen changes of visual stimuli [11]. An important contribution from this field of research has been introducing synchronization methodologies for combining eye tracking with EEG to determine the exact time at which the user initiates a response to a given stimulus (using fixation-locking) rather than simply relying on timing information for when a given stimulus was presented. This is especially useful for understanding how people interact with LLMs (ChaGPT, Gemini, Claude, LlaMA, DeepSeek) including the times when they decide to read something again, verify it, and when they shift focus. A key aspect of multi-modal reasoning in Neuro-Information Systems (Neuro-IS), then, is that neurophysiologic techniques may be able to identify and measure cognitive mechanisms which are currently unmeasured or at least not measured as effectively by using behavioral measures [2]. More recently, recent studies on Neuro-IS have also highlighted the need for ecological validity of measures such that physiological measures are aligned with realistic interaction scenarios [6]. Thus, we utilize both EEG and Eye-tracking, along with manipulation checks and short self-reports, while utilizing an initial positive-neutral-negative (PNN) affect calibration block to support interpretation of LLM [4].

Hallucinations and attention behavior

Computational linguistics research shows that human eye movements provide useful data for hallucination search and can be used to develop hallucination-detecting models [7]. This leads to the belief that measurable changes to the ways students read and verify information will occur as a result of hallucination exposure (longer time spent looking at questionable statements, greater number of backtracking actions (regressions), and wider area of source scanned. Importantly, perception is not guaranteed: some hallucinations may go unnoticed, while others may trigger rapid uncertainty and verification effort. This design distinguishes objective hallucination exposure from perceived hallucination using trial-level manipulation checks, enabling us to test whether attention effects depend on detection of incorrectness [5].

EEG and eye tracking as a viable classification approach

Previous studies conducted on using both EEG and eye tracking has indicated that using joint EEG and eyes as biomarkers could be used for classification with over a 60% success rate in early pilot studies [8]. Thus, supporting our plan to test the use of ensemble learning methods (Random Forest & XGBoost) vs. simple baseline methods (Decision Trees) when classifying emotional responses during hallucinations. In the revised framing, machine learning is positioned as an analytical tool for inference rather than the theoretical contribution. The core contribution is the multi-level mechanism linking hallucination exposure to affective appraisal, attention behavior, and performance, consistent with NeuroIS research agendas [2][5].

Research Model and Hypothesis

Conceptual Model

Hallucination episodes are modeled as interactive events that evoke specific affective responses (anxiety, frustration, uncertainty) and attentional behaviors (re-reading, scanning, verification efforts) that contribute to performance (accuracy, task completion time, answer quality). The influence of individual characteristics (trust propensity, LLM experience, AI literacy) is also expected to mediate these relationships. To address construct clarity, the revised model distinguishes objective hallucination exposure (manipulated hallucinated vs matched accurate outputs) from perceived hallucination (measured via trial-level manipulation checks). Conceptually, hallucination encounters function as technology-mediated affective events that trigger rapid appraisal (perceived reliability) and verification behaviors. EEG and eye tracking are used as indicators of underlying arousal and attention allocation, and are triangulated with brief self-report measures rather than treated as direct readouts of discrete emotions (Fig. 1).

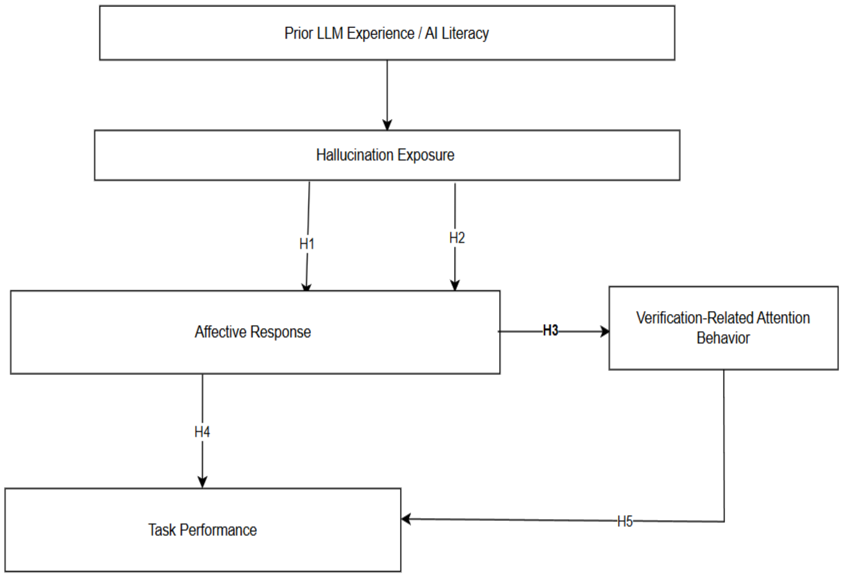


Fig 1. Methodological Model (Author's own work)

The key concepts included in this study include, hallucination exposure (exposure to an LLM output including incorrect or unsupported information), affective response (the emotional state experienced by a student immediately before and after a hallucination episode), behavioral attention (eye movement data and fixation patterns and verification behaviors derived from eye movement data using eye-tracking), and student performance (accuracy and quality of answers submitted and efficiency-based metrics).

Hypothesis

H1: Exposure to hallucinations will increase negative affective states compared to interaction segments without hallucinations (as observed through both EEG and eye metrics).
H2: Exposure to hallucinations will be associated with greater verification-related attention patterns (longer fixation times and regression times on relevant content and greater gaze dispersion indicative of search and uncertainty).
H3: Affective response states exhibited during hallucination episodes will negatively correlate with performance (lower accuracy, quality, and longer completion times)
H4: The inclusion of EEG and eye-tracking features in models will allow for better classification of emotional response states and prediction of performance than models utilizing EEG-only or eye-only features.

H5: Students' Prior LLM experience weakens the negative impact of hallucination exposure on perceived frustration and attenuates the negative association between hallucination-triggered affective response and performance by enabling more efficient verification strategies.

Experimental Method (Work in Progress)

Participants and status

We are collecting the first simultaneous EEG and eye-tracking data from 32 students (ongoing), to strengthen internal validity for affect labeling and improve interpretability of physiological signals, we include a brief affect calibration block using a positive–neutral–negative (PNN) images/videos framework prior to the LLM task trials [9].

Task and experimental flow

Students perform an academic task on an LLM interface (answer a question in a subject area or create a written response). To control for hallucinations, we have manipulated controlled output of the LLM by providing participants with either LLM outputs that include pre-seeded hallucination segments versus outputs that are accurate and matched. The time stamps of task events and the times-tamps of responses to the LLM will be logged to allow us to precisely define windows for the extraction of both EEG and eye-tracking features (Fig. 2) (baseline, exposure, response, verification) [5].

Fig. 2. Eye Tracking measurement of Eye Movements, Gaze Data Chart, Pupil Diameter, and Eye Openness (Author's own work)

Measurements

Eye tracking (Tobii), fixation count, gaze dispersion, AOI-based metrics (region of claim vs region of evidence), saccadic features, pupil-related features as applicable. Eye tracking lends itself well to a quick look at how the participant visually inspects the material and how their fixation behaviors change. For EEG, frequency band-based features (band-power delta/theta/alpha/beta/gamma) obtained through spectral analysis (consistent with the typical method of analyzing EEG data). Low-intrusive affect measures at designated check points (short valence) to provide grounding for emotional state labels and to enhance the diagnostic utility of physiological measures (physiological measures are often associated with more than one construct, thus, when possible, use a triadic approach to increase the interpretability of results).

Fig.3. EEG Frequency Measurement (Author's own work)

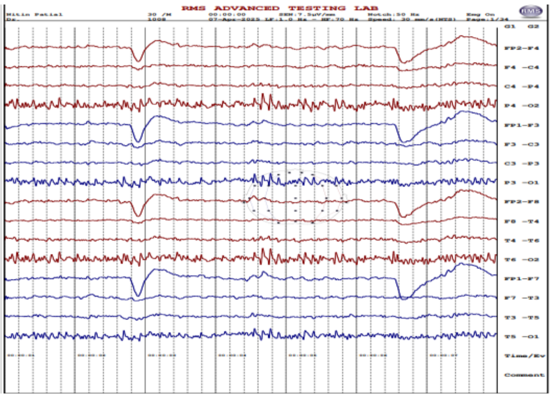


Fig.4. EEG Frequency Measurement Results (Author's own work)

Data Processing and Analysis Plan

Preprocessing and synchronization

For Eye tracking, blink and missing data removal, AOI mapping, fixation parsing. EEG is to be filtered, artifacts removed, segmented by time windows into episodes, and multimodal synchronization of the streams with a common timestamp. In psychophysiology, synchronization is important to maintain both objectivity and interpretability in the study of physiological processes that occur in conjunction with psychological processes [5]. Episodes are segmented into baseline → hallucination/accurate exposure → verification window.

Feature extraction

EEG will be used as an input to create features based on relative band power, engagement, and event related amplitudes. Frequency analysis and ERP methods are commonly employed in the literature to identify discrete event elicited cognitive and affective responses from EEG data. Eye tracking will be used to produce features based on mean fixation duration, fixation count, gaze dispersion, AOI dwell times, transitions between AOIs, and pupil derived arousal proxy values. EEG and pupil-based features are interpreted primarily as markers of arousal and cognitive control demand.

Modeling strategy

Supervised learning models will be evaluated to predict emotional response states and to predict performance. Decision Tree (interpretable baseline), Random Forest (feature importance for interpretability, robust ensemble model), and XGBoost (gradient-boosted trees, high performance on tabular features) will be applied.

Expected Contributions and Implications

Theoretical contribution

The current research extends existing literature in NeuroIS by examining how emotional responses are triggered through hallucinations within LLM mediated learning, providing evidence that these hallucinations create rapid changes in an individual's emotional state which ultimately impact their processing of information and performance. Using NeuroIS methods to capture emotions while they occur provides a significant advantage over self-report methods alone.

Method contribution

This study develops and validates a replicable, episode-based protocol to assess LLM interactions using both EEG and eye tracking simultaneously, and addresses some of the specific methodological calls in NeuroIS to ensure measurement validity, diagnosticity, and objectivity. This method has particular value in conversational AI systems since users have control over the amount of time spent reading and verifying information and thus fixation-locked or event-based analyses are appropriate. EEG and pupil-based features are interpreted primarily as markers of arousal and cognitive control demand.

Design and practical implications

Findings from this study will be useful in designing educational LLM interfaces that are emotionally aware and safe for students. Examples include providing cues to indicate when there may be uncertainty or whether the response is grounded when the risk of hallucination is high, prompting the student to verify information when the gaze pattern suggests that the student is only superficially scanning, and providing the student with supporting language when the neurophysiological data indicates that the student's negative affective state is increasing.

References

1. Lutz, B., Adam, M.T.P., Feuerriegel, S., Pröllochs, N., Neumann, D.: Lutz, B., Adam, M.T.P., Feuerriegel, S., Pröllochs, N., Neumann, D.: Affective Information Processing of Fake News: Evidence from NeuroIS. In: Davis, F.D., Riedl, R., vom Brocke, J., Léger, P.M., Randolph, A.B. (eds.) Information Systems and Neuroscience. Springer (updated edition/year per final bib).
2. Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Mueller-Putz, G.R., Pavlou, P.A., Riedl, R., vom Brocke, J., Weber, B.: On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS Quarterly 36(3), 679–702 (2012).
3. Riedl, R., Davis, F.D., Hevner, A.R.: Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. J. Assoc. Inf. Syst. 15(10), 1–35 (2014).
4. Riedl, R., Léger, P.-M.: Fundamentals of NeuroIS: Information Systems and the Brain. Springer (2016).
5. vom Brocke, J., Hevner, A., Léger, P.-M., Walla, P., Riedl, R.: Advancing a NeuroIS research agenda with four areas of societal contributions. Eur. J. Inf. Syst. 29(1), 9–24 (2020).
6. Balapour, A., Riedl, R.: Ecological validity in NeuroIS research: Theory, evidence, and a roadmap for future studies. Journal of the Association for Information Systems (2025).
7. Anderson, B.B., Vance, A., Kirwan, C.B., Eargle, D., Jenkins, J.L.: How users perceive and respond to security messages: a NeuroIS research agenda and empirical study. Eur. J. Inf. Syst. 25, 364–390 (2016).
8. Nagendran, M., Festor, P., Komorowski, M., Gordon, A., Faisal, A.A.: Eye-tracking of clinician behaviour with explainable AI decision support: a high-fidelity simulation study. In: ICML Workshop (2023).
9. Zhang, G., Cai, Z., Shi, J., He, J., Li, J., Chen, L.: Integrating LLM, EEG, and eye-tracking biomarker analysis for word-level neural state classification in semantic inference reading comprehension. In: IEEE BIBM (2024).
10. Maharaj, K., Saxena, A., Kumar, R., Mishra, A., Bhattacharyya, P.: Eyes show the way: Modelling gaze behaviour for hallucination detection. Findings of ACL: EMNLP (2023).
11. Zhang, D.: Towards attention-aware large language models: integrating real-time eye-tracking and EEG for adaptive AI responses. arXiv (2025).

Investigating Reliance on Artificial Intelligence Through Adaptive Eye-Tracking: A Conceptual Paper and Research Agenda

Luis Gutmann[1,*], Christiane Ernst[1], Aneta Mertová [1], and Kathrin Figl[1]

[1] University of Innsbruck, Institute of Information Systems, Production and Logistics, Innsbruck, Austria
{luis.gutmann,christiane.ernst,aneta.mertova,kathrin.figl}@uibk.ac.at

Abstract. Adaptive eye-tracking transforms attention from a retrospective measure into an active control signal, allowing digital systems to dynamically adjust to real-time gaze. While increasingly applied in psychology, marketing, and human-computer interaction, its potential for studying human reliance on artificial intelligence (AI) in information systems (IS) research remains largely unexplored. This paper introduces a conceptual framework and a practical methodological implementation for adaptive eye-tracking research, realized through a modular setup combining Tobii eye trackers, PsychoPy, and the Titta toolbox. By enabling gaze-contingent experimental designs, adaptive eye-tracking allows researchers to investigate AI reliance at the process level through real-time measures of attentional allocation, verification behavior, and cognitive load, positioning it as a promising methodological approach for NeuroIS research on human-AI interaction.

Keywords: Adaptive Eye-tracking • Real-Time Adaptation • NeuroIS Methods • AI Reliance

Introduction

Artificial Intelligence (AI) increasingly shapes human decision-making across contexts such as recommender systems and generative AI tools and decision support systems. In many of these settings, users must decide when, and to what extent to rely on algorithmic advice. While information systems (IS) research has examined AI reliance primarily through behavioral outcomes and self-reports [e.g, 1, 2], the real-time attentional and cognitive processes that precede reliance decisions remain underexplored [3]. Addressing this gap requires methods that capture how users engage with AI-generated information as the interaction unfolds. Eye-tracking, traditionally used in IS research as a diagnostic method to retrospectively analyze attention and decision processes, offers such potential. Recently, it can also be employed adaptively, treating gaze as a real-time input that dynamically modifies system behavior. This enables a dynamic interplay, allowing the system to respond to users' attention while the human-AI interaction is still ongoing.

Despite its methodological potential, adaptive eye-tracking has not been systematically integrated into IS research on AI reliance, even though prior reliance research has employed eye-tracking diagnostically [e.g., 4, 5, 6, 7]. Accordingly, this paper addresses the following research question: *How can adaptive eye-tracking be conceptualized and implemented to investigate reliance on AI in IS and NeuroIS research?* To answer this, we develop a conceptual framework, outline a research agenda to investigate AI reliance using real-time gaze-contingent experimental designs, and introduce an implementable methodological setup.

Concept and Foundations

Eye-tracking is well established in IS as a research method [8]. Traditionally, it has been used diagnostically: gaze behavior is recorded during interaction for retrospective process tracing to analyze interface design or decision-making without influencing the system itself [8, 9]. A key distinction exists between diagnostic and interactive eye-tracking [9]. While diagnostic use treats gaze as an outcome variable, interactive use conceptualizes gaze metrics as a real-time input channel. This shift from passive measurement to active control motivates adaptive eye-tracking, where gaze modifies system behavior during interaction.

Adaptive eye-tracking builds on adaptive systems research, which defines adaptation as modifying system behavior based on processed user information [10]. Early biocybernetic systems showed that psychophysiological signals can regulate human-machine interaction [11]. In contrast to survey-based personalization or clickstream-driven recommender systems [12], adaptive eye-tracking uses ocular metrics, including fixations, dwell times, gaze transitions, and pupillary responses, as real-time neurophysiological triggers within continuous control loops. Gaze thus functions as a design input that reflects attention and cognitive processing [13] rather than merely an outcome variable [14]. Adaptive eye-tracking is conceptualized as a physio-adaptive closed-loop architecture that integrates gaze acquisition, real-time processing, trigger logic, and stimulus modification. Simple fixation- or dwell-based thresholds operationalize attention as a control parameter [15, 16, 17]. More advanced systems incorporate pupil responses as indicators of users' cognitive states, such as mental effort or cognitive load [18], enabling cognitive load-sensitive adaptation in immersive and high-stakes environments [19]. Recent extensions combine gaze with EEG or generative AI-based inference models to enhance contextual precision [20, 21].

Methodologically, these systems often operationalize attention using gaze metrics derived from predefined Areas of Interest (AOIs), assuming that fixations within these regions indicate task-relevant processing [22]. Gaze-based adaptations have been studied in web environments [12], gaze-sensitive interfaces [16], and cognitive load-adaptive systems [23], including VR-based training systems [19]. Research has evolved from inferring metacognitive skills [24] to multimodal and AI-driven real-time adaptation.

Real-world applications of adaptive eye-tracking span diverse domains, including medical training [19], purchasing decisions [25], product recommendation [26], adaptive online advertisements and news presentation [12], assistive technologies for individuals with special needs [27], adaptive reading displays [28], map and aerial image interaction [20], gaming and learning [27], and aviation contexts enhancing attention guidance and workload assessment [29]. Integration with digital nudges has also been proposed to increase intervention effectiveness [30]. Taken together, these research streams show how systems can sense attention and respond in ways that directly shape interaction and decision-making.

In sum, adaptive eye-tracking marks a transition from retrospective process tracing to real-time, attention-aware system design. Grounded in adaptive systems theory and the diagnostic-interactive distinction, it integrates neurophysiological measures directly into system behavior. This promise comes with practical constraints: synchronized acquisition, low-latency processing, and reliable adaptation trigger logic are prerequisites for valid gaze-contingent research.

Research Agenda: Investigating Reliance on AI through Real-Time Gaze-Contingent Experimental Designs

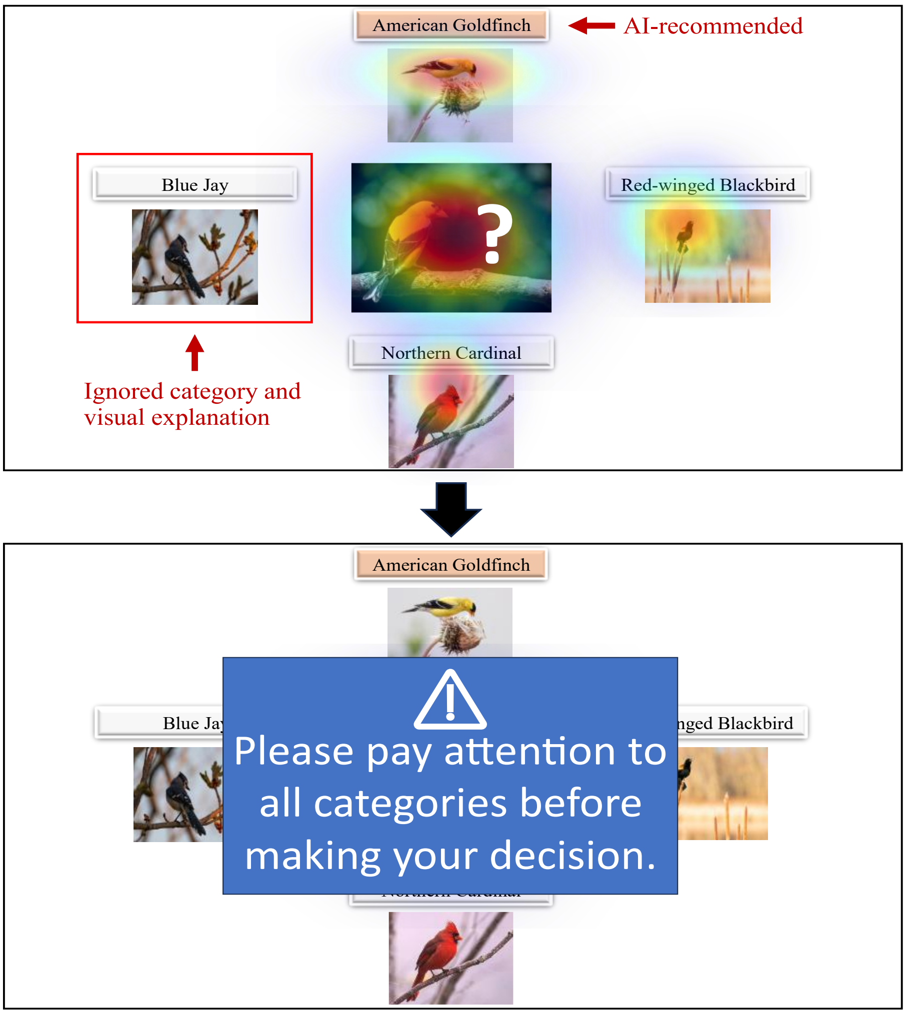
Reliance on AI systems is typically inferred from observable outcomes such as agreement rates [31], weight of advice [32], task performance [33], or self-reported trust [34]. While informative, these measures leave a critical blind spot: they do not reveal whether users actually attend to, process, or evaluate the information provided by the system. Real-time gaze-contingent experimental designs offer a way to disentangle reliance, automation bias and superficial compliance by operationalizing attention as a measurable and adaptive mediator of reliance formation.

We propose reframing AI reliance research around three core information-processing mechanisms: attentional allocation, verification behavior, and cognitive load. Attention functions as a scarce cognitive resource that mediates the relationship between AI interface design and reliance outcomes. Many central phenomena in the AI literature - including algorithm aversion [35, 36] and appreciation [37] - may stem not from flawed attitudes but from selective or insufficient inspection of model outputs, uncertainty cues, or explanations. By capturing fixation patterns, dwell times, scanpaths, and pupil dynamics, adaptive eye-tracking enables process-level modeling of how users engage with AI-generated information before deciding to accept or override it, and enables gaze-contingent adaptations based on users' real-time attention.

Explainable AI (XAI) provides a particularly suitable testbed for this agenda. Existing research frequently assumes that explanations improve reliance, yet rarely examines whether users actually inspect those explanations [1], [38, 39]. A gaze-contingent experimental design can differentiate between (1) users who ignore explanations, (2) users who inspect explanations but still reject the AI recommendations, and (3) users who deeply process explanatory content. For instance, in an image-classification task (e.g., as applied in [39, 40]) where participants evaluate AI diagnoses or object labels, gaze data can reveal whether visual heatmaps, uncertainty indicators, or confidence scores receive sustained attention. Adaptive mechanisms can then be introduced: if explanation areas remain uninspected, the system may increase saliency, introduce brief verification prompts, or temporarily delay acceptance of an AI recommendation. Crucially, this allows testing whether explanation effects depend on attentional engagement and whether interventions reduce automation bias while preserving efficiency. Fig. 19 illustrates this principle using a bird-classification task (also used, for example, in [39]), where the system issues a warning if a category or its corresponding visual explanation has not received sufficient attention - a design we implemented and validated as described in Section 4.

Fig. 19. Gaze-triggered adaptive warning for unattended example-based explanation

Methodologically, gaze-contingent paradigms strengthen reliance research in three



ways. First, they enable temporally precise reconstruction of information acquisition sequences, allowing researchers to trace how users process AI-generated information before forming reliance decisions. Second, objective gaze metrics reduce construct ambiguity by distinguishing non-inspection from informed rejection. Third, integrating attention into experimental designs enhances mediation analysis, clarifying whether design interventions influence reliance directly or through cognitive engagement.

Finally, attention-aware AI systems raise ethical and governance questions when transferred from research settings with participants' informed consent to real-world applications [47]. Systems that detect whether users inspect explanations can infer aspects of attention and cognitive processing, thereby increasing their potential to shape user judgments and behavior. Transparency, consent, and limits on adaptive influence therefore become central design considerations. By embedding attention as a measurable and adaptive construct, IS research can move from outcome-based assessments of reliance toward a process-sensitive, theoretically grounded understanding of human-AI interaction. The following section presents a concrete methodological foundation for realizing this agenda, introducing a modular framework for implementing adaptive eye-tracking in NeuroIS research.

Proposed Methodological Framework

Adaptive eye-tracking requires the integration of experimental control, real-time gaze acquisition, and precise temporal synchronization. Implementation approaches range from low-level programming via manufacturer software development kits (SDKs) to proprietary integration environments. Direct SDK integration offers flexibility but requires substantial engineering effort. We therefore adopt a modular and accessible solution by combining PsychoPy, a Tobii eye tracker, and the Titta toolbox [48, 49]. This framework enables real-time gaze-contingent experimental designs without custom low-level programming while maintaining full access to live gaze streams, synchronization mechanisms, and structured data storage. Together, these components form a closed-loop system that enables millisecond-level, real-time adaptation.

System Components of the Closed-Loop Architecture

The architecture comprises four functionally distinct yet tightly integrated components: (1) gaze data acquisition, (2) real-time processing, (3) experimental and adaptive stimulus control, and (4) post-hoc analysis. This ordering reflects the data flow within the closed-loop architecture (see Fig. 20).

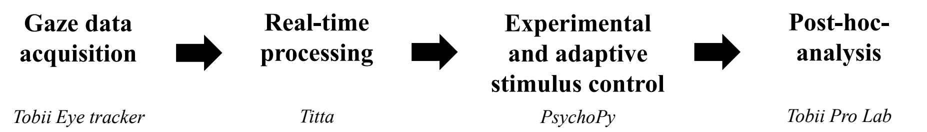


Fig. 20. Closed-loop control architecture

The Tobii eye tracker constitutes the gaze data acquisition layer. Crucially, data acquisition runs independently of the experimental loop, ensuring that temporal precision is not affected by stimulus rendering or processing overhead in PsychoPy. This independence stabilizes the gaze data stream and ensures consistent sampling intervals regardless of experimental complexity.

Titta serves as the Tobii-PsychoPy integration layer, specifically designed to provide full access to Tobii eye tracker capabilities within PsychoPy. Technically, Titta relies on TittaPy, a C++ wrapper around the Tobii SDK, to minimize the risk of missing samples during continuous gaze streaming. For researchers, Titta provides: (1) device initialization using model-specific default parameters; (2) an intuitive, built-in calibration and validation interface embedded within the PsychoPy interface; (3) a continuous gaze buffer that streams and stores incoming samples independently of the experimental loop; and (4) timestamp synchronization between gaze data and experimental events [48].

PsychoPy serves as the experimental control layer. As an open-source platform widely used for behavioral and cognitive experiments, PsychoPy manages stimulus presentation and trial flow, executes adaptive decision rules, and logs all experimental events and data. All adaptation logic is implemented within PsychoPy. This centralization of experimental logic ensures that gaze-contingent adaptations remain fully documented and replicable across research contexts.

Tobii Pro Lab extends the architecture with post-hoc analysis capabilities. Via the TalkToProLab class, AOIs and stimulus events defined in PsychoPy can be automatically registered and mirrored in Tobii Pro Lab, enabling direct comparison between the real-time adaptive logic and post-hoc gaze analysis. This integration is particularly valuable for validation purposes, as it allows researchers to verify that real-time adaptive decisions correspond to post-hoc gaze analyses within the same environment.

The system follows a layered communication model that ensures temporally precise interaction between experimental logic and gaze acquisition. Initially, a connection to the Tobii device is established through Titta, which initializes the eye tracker. Recording is initiated only after calibration quality meets predefined spatial accuracy thresholds. During execution, gaze samples are continuously streamed and buffered via Titta, enabling PsychoPy to access them in real time for adaptive rule implementation. When a predefined trigger condition is satisfied, PsychoPy modifies stimulus properties or alters trial flow accordingly, while preserving precise timestamp synchronization between gaze signals and experimental events.

Proof-of-Concept Design

To validate the proposed architecture, we implemented a simple gaze-contingent experimental paradigm and conducted an initial author-led validation in a controlled laboratory environment. The primary goal of this validation was the technical verification of the closed-loop logic, given its proof-of-concept nature, rather than behavioral experimentation with external participants or the development of a fully specified experimental design. The experiment was developed in PsychoPy 2022.1.3 on Windows 11, using a Tobii Pro Spectrum eye tracker with firmware version 2.6.1, and a dual-screen setup: one screen served as the experimenter monitor, while the participant-facing screen presented the experimental stimuli.

Task Logic. The task consisted of a bird classification paradigm designed as an example-based XAI task, which was introduced as a conceptual example in Section 3. A target bird was displayed in the center of the screen and surrounded by candidate species that the AI model identified as the most likely options. Each candidate species was accompanied by a representative example image and the corresponding species name (see Fig. 19). The goal was to correctly assign the target bird to one of the displayed candidate species based on visual inspection. In this way, the task presented participants with a set of plausible AI-generated options, while the example images served as example-based explanations supporting their decision-making.

Two-Page Adaptive Paradigm. The experiment consisted of two stimulus pages. The first page presented the bird classification task as described above, with AOIs set on each of the four candidate birds and their corresponding species labels. The system continuously evaluated cumulative dwell time per AOI during the observation period. If the inspection threshold was met for all AOIs individually, the experiment concluded without intervention. If one or more AOIs remained below the threshold, the second page triggered an adaptive warning prompting the observer to inspect all candidates before reaching a decision. Importantly, the warning did not reveal which specific bird had been insufficiently inspected, preserving the integrity of the classification decision. Further details of the experimental design were not elaborated, given the proof-of-concept nature of the validation. The primary objective was to verify the technical implementation of the adaptive trigger mechanism, rather than to optimize the behavioral intervention or to develop a fully specified experimental design.

Gaze-Contingent Trigger Implementation. AOIs were defined in PsychoPy in screen pixel coordinates for each of the four candidate species, covering both their representative example images and corresponding species labels. During the observation period, gaze data were continuously streamed from the Tobii eye tracker, buffered through Titta, and passed to PsychoPy in real time. For each valid gaze sample, PsychoPy evaluated whether the gaze coordinates fell within a predefined AOI and accumulated the corresponding dwell time. At trial end, the cumulative dwell time per AOI was compared individually against a predefined threshold. Based on this comparison, PsychoPy either concluded the experiment directly - if all AOIs met the threshold - or triggered the adaptive warning page if one or more AOIs fell below it. This simple adaptive design was deliberately chosen to isolate and validate the real-time gaze classification, dwell time accumulation, and gaze-contingent stimulus selection.

Technical Validation and Implementation Challenges

Validation. Validation focused on confirming correspondence between the real-time adaptive logic and post-hoc Tobii Pro Lab analysis across three dimensions. First, AOI coordinates defined in PsychoPy were verified against those registered in Tobii Pro Lab, both visually - by overlaying AOI boundaries on the stimulus during the experiment - and through sample-level comparison, confirming exact positional agreement. Second, dwell time computations generated by PsychoPy during the experiment were compared against post-hoc fixation metrics in Tobii Pro Lab, with no discrepancies observed between real-time output and post-hoc analysis. Third, adaptive trigger decisions based on real-time gaze classification were cross-checked against post-hoc AOI hit data in Tobii Pro Lab, again revealing no discrepancies. Formal latency measurement was outside the scope of this proof-of-concept. Nevertheless, no noticeable delays in adaptive responses were observed during testing, which is plausible given the computational simplicity of the threshold-based trigger logic. More complex experiments with longer durations may warrant additional consideration of storage and synchronization time. Future implementations involving more complex trigger logic - such as pupil-based cognitive load estimation - should empirically assess temporal precision.

Implementation Challenges and Practical Recommendations. Several implementation challenges arose during the proof-of-concept that are worth documenting for researchers adopting this framework.

Firmware and driver compatibility proved to be a critical prerequisite. In our setup, the Tobii Pro Spectrum was initially running firmware version 2.9.0-oculus-0 - a variant intended for Oculus integration - which caused the device not to be recognized as a standard Tobii Pro eye tracker. The firmware was reverted using the Tobii Eye Tracker Manager, after which the device was correctly recognized. Researchers adopting this framework should ensure version compatibility across all components (see Section Proof-of-Concept Design).

Connection initialization between PsychoPy and Tobii Pro Lab represented a second challenge. The TalkToProLab connection is state-dependent: when PsychoPy initializes the experiment, Tobii Pro Lab must already have an active External Presenter project open with the record tab active. If these conditions are not met at initialization, both PsychoPy and Tobii Pro Lab crash, requiring both applications to be restarted before the experiment can be launched again. We therefore recommend verifying Pro Lab's connection state explicitly before experiment initialization to avoid this failure mode.

Stimulus event management in Tobii Pro Lab introduced a third challenge. When managing stimulus events directly through PsychoPy and transferring data to Tobii Pro Lab, each stimulus event must be explicitly closed by logging both a start and an end timestamp - a requirement that is not documented in the Titta documentation. In our implementation, data transfer from PsychoPy to Tobii Pro Lab repeatedly failed after experiment completion because stimulus events were not explicitly closed, causing Tobii Pro Lab to remain in the state "awaiting finalization of recording." Ensuring that every stimulus event was properly closed before finalizing the recording resolved this issue. Researchers adopting this framework should therefore account for explicit event closure as a prerequisite for successful data integration between PsychoPy and Tobii Pro Lab.

Conclusion

AI reliance research in IS has largely focused on outcomes while leaving the underlying attentional processes unobserved. This paper argues that adaptive eye-tracking enables process-level analysis by incorporating real-time gaze into human-AI interaction, allowing researchers to distinguish reliance based on actual information processing versus reliance without inspection. By treating attention as a measurable and adaptive construct, we offer a conceptual and practical foundation for researching how AI reliance forms - and how it can be shaped - during interaction. A key contribution lies in demonstrating that adaptive eye-tracking is not only theoretically promising but also practically accessible: the proposed framework combines PsychoPy, Tobii eye trackers, and the Titta toolbox into a modular solution that does not require extensive low-level programming expertise, lowering the barrier for IS and NeuroIS researchers to adopt gaze-contingent experimental designs.

Beyond XAI, the adaptive eye-tracking approach can be applied to misinformation detection and online persuasion [41, 42, 43], (clinical) decision support [44, 45], or forecasting [46], all application areas currently discussed in IS research. In these contexts, errors often arise not because relevant information is unavailable, but because it is unattended. Real-time attention modeling enables systems to detect shallow processing, attentional tunneling, or overload, and to adapt interface complexity accordingly. Building on the validated proof-of-concept, the next step is the design and execution of a first full-scale experiment with external participants, investigating AI reliance through real-time gaze-contingent designs. Future work should also explore more advanced trigger logic, such as pupil-based cognitive load estimation and multimodal combinations of gaze metrics, to further extend the framework's applicability across diverse research contexts.

References

1. Eckhardt, S., Kühl, N., Dolata, M., Schwabe, G.: A Survey of AI Reliance. ACM Comput. Surv. (2025)
2. You, S., Yang, C.L., Li, X.: Algorithmic versus Human Advice: Does Presenting Prediction Performance Matter for Algorithm Appreciation? Journal of Management Information Systems 39, 336-365 (2022)
3. Litvinova, Y., Keplinger, K.: XAI-Supported Decision-Making: Insights from NeuroIS Studies for a User Perspective. NeuroIS Retreat, pp. 157-177. Springer (2024)
4. Kim, J.Y., Marshall, V.D., Rowell, B., Chen, Q., Zheng, Y., Lee, J.D., Kontar, R.A., Lester, C., Yang, X.J.: The effects of presenting AI uncertainty information on pharmacists' trust in automated pill recognition technology: exploratory mixed subjects study. JMIR Human Factors 12, e60273 (2025)
5. Liu, J., Marriott, K., Dwyer, T., Tack, G.: Increasing User Trust in Optimisation through Feedback and Interaction. ACM Trans. Comput.-Hum. Interact. 29, Article 42 (2023)
6. Manresa-Yee, C., Ramis, S., Gaya-Morey, F.X., Buades, J.M.: Impact of explanations for trustworthy and transparent artificial intelligence. Proceedings of the XXIII International Conference on Human Computer Interaction, pp. 1-8 (2023)
7. Tadson, B., Krisam, C., Mädche, A.: Understanding Overreliance in Human-GenAI Interactions and Its Impact on Divergent Thinking: A NeuroIS Study. NeuroIS Retreat, pp. 143-157. Springer (2025)
8. Vasseur, A., Passalacqua, M., Sénécal, S., Léger, P.-M.: The use of eye-tracking in information systems research: a literature review of the last decade. AIS Transactions on Human-Computer Interaction 15, 292-321 (2023)
9. Duchowski, A.: Eye tracking methodology: Theory and practice. Springer-Verlag (2007)

10. Oppermann, R.: From user-adaptive to context-adaptive information systems (von benutzeradaptiven zu kontextadaptiven informationssystemen). I-com 4, 4-14 (2005)
11. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biological psychology 40, 187-195 (1995)
12. Alt, F., Shirazi, A.S., Schmidt, A., Mennenöh, J.: Increasing the user's attention on the web: using implicit interaction based on gaze behavior to tailor content. Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, pp. 544–553. Association for Computing Machinery, Copenhagen, Denmark (2012)
13. Shahriar, A.B., Moon, M., Mahmud, H., Hasan, K.: Online Product Recommendation System by Using Eye Gaze Data. ACM Symposium on Eye Tracking Research and Applications, pp. 2020 (2020)
14. Zhao, Q., Chang, S., Harper, F.M., Konstan, J.A.: Gaze prediction for recommender systems. Proceedings of the 10th ACM Conference on Recommender Systems, pp. 131-138 (2016)
15. Toreini, P., Langner, M., Maedche, A., Morana, S., Vogel, T.: Designing Attentive Information Dashboards. Journal of the Association for Information Systems 23, 521-552 (2022)
16. Menges, R., Kumar, C., Staab, S.: Improving User Experience of Eye Tracking-Based Interaction: Introspecting and Adapting Interfaces. ACM Trans. Comput.-Hum. Interact. 26, Article 37 (2019)
17. Xu, S., Jiang, H., Lau, F.C.M.: Personalized online document, image and video recommendation via commodity eye-tracking. Proceedings of the 2008 ACM conference on Recommender systems, pp. 83-90 (2008)
18. Dimoka, A., Davis, F. D., Gupta, Banker, R. D., Dennis, A. R., Ischebeck, Müller-Putz, G. R., Benbasat, I., Gefen, et al.: On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. MIS Quarterly 36, 679-702 (2012)
19. Funk, M.-P., Navab, N., Schütz, L.: Eye Tracking-Based Adaptive User Interfaces in Virtual Reality Eye Surgery Training. Proceedings of the 2025 Mensch und Computer 2025, pp. 578-582 (2025)
20. Qin, T., Fias, W., Van de Weghe, N., Huang, H.: Recognition of map activities using eye tracking and EEG data. International Journal of Geographical Information Science 38, 550-576 (2024)
21. Schulz, T., Krisam, C., Seitz, J.: EyeGPT - A Cognitive Load-Adaptive GenAI Assistant with Eye Tracking for Programming Education. NeuroIS, (2025)
22. Serim, B., Chech, L., Vasilieva, M., Papa, L., Gamberini, L., Jacucci, G.: Exploring gaze-adaptive features for interacting with multi-document visualizations. Proceedings of the 10th International Symposium on Visual Information Communication and Interaction, pp. 85–92. Association for Computing Machinery, Bangkok, Thailand (2017)
23. Gebhardt, C., Hecox, B., Van Opheusden, B., Wigdor, D., Hillis, J., Hilliges, O., Benko, H.: Learning cooperative personalized policies from gaze data.

Proceedings of the 32nd annual ACM symposium on user interface software and technology, pp. 197-208 (2019)
24. Conati, C., Merten, C., Amershi, S., Muldner, K.: Using Eye-Tracking Data for High-Level User Modeling in Adaptive Interfaces. AAAI Conference on Artificial Intelligence, (2007)
25. Feit, A.M., Vordemann, L., Park, S., Berube, C., Hilliges, O.: Detecting Relevance during Decision-Making from Eye Movements for UI Adaptation. ACM Symposium on Eye Tracking Research and Applications, pp. 1-11 (2020)
26. Cheng, S., Liu, X., Yan, P., Zhou, J., Sun, S.: Adaptive user interface of product recommendation based on eye-tracking. Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction, pp. 94–101. Association for Computing Machinery, Hong Kong, China (2010)
27. Chahine, W., Hachem, N., Moacdieh, N.M.: Eye Tracking-Based Adaptive Displays: A Review of the Recent Literature. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 67, 927-932 (2023)
28. Iqbal, S.T., Bailey, B.P.: Using eye gaze patterns to identify user tasks. The Grace Hopper Celebration of Women in Computing, vol. 4, pp. 2004. Illinois, USA (2004)
29. Xenos, M., Mallas, A., Minas, D.: Using Eye-Tracking for Adaptive Human-Machine Interfaces for Pilots: A Literature Review and Sample Cases. Journal of Physics: Conference Series 2716, (2024)
30. Hummel, D., Toreini, P., Maedche, A.: Improving Digital Nudging Using Attentive User Interfaces: Theory Development and Experiment Design. Designing the Digital Transformation: DESRIST 2018 Research in Progress Proceedings of the 13th International Conference on Design Science Research in Information Systems and Technology. Chennai, India. 3 Jun-6 Jun, (2018)
31. He, G., Aishwarya, N., Gadiraju, U.: Is conversational xai all you need? human-ai decision making with a conversational xai assistant. Proceedings of the 30th International Conference on Intelligent User Interfaces, pp. 907-924 (2025)
32. Mahmud, H., Islam, N., Krishnan, S.: Human-Robo-advisor collaboration in decision-making: Evidence from a multiphase mixed methods experimental study. Decision Support Systems 198, 114541 (2025)
33. Bockstedt, J.C., Buckman, J.R.: Humans' use of ai assistance: The effect of loss aversion on willingness to delegate decisions. Management Science 72, 323-342 (2026)
34. Bertrand, A., Eagan, J.R., Maxwell, W.: Questioning the ability of feature-based explanations to empower non-experts in robo-advised financial decision-making. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pp. 943–958. Association for Computing Machinery, Chicago, IL, USA (2023)
35. Dietvorst, B.J., Simmons, J.P., Massey, C.: Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. Journal of Experimental Psychology: General 144, 114 (2015)

36. Jussupow, E., Benbasat, I., Heinzl, A.: An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making. MIS Quarterly 48, 1575-1590 (2024)
37. Logg, J.M., Minson, J.A., Moore, D.A.: Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. Organizational Behavior and Human Decision Processes 151, 90-103 (2019)
38. Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M.S., Krishna, R.: Explanations can reduce overreliance on ai systems during decision-making. Proceedings of the ACM on Human-Computer Interaction 7, 1-38 (2023)
39. Morrison, K., Spitzer, P., Turri, V., Feng, M., Kühl, N., Perer, A.: The impact of imperfect XAI on human-AI decision-making. Proceedings of the ACM on human-computer interaction 8, 1-39 (2024)
40. Fügener, A., Grahl, J., Gupta, A., Ketter, W.: Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. Information Systems Research 33, 678-696 (2022)
41. Baiyere, A., Bauer, J.M., Constantiou, I., Hardt, D.: Fake News and True News Assessment: The Persuasive Effect of Discursive Evidence in Judging Veracity. MIS Quarterly 49, 823-860 (2025)
42. Moravec, P.L., Minas, R.K., Dennis, A.R.: Fake News on Social Media. MIS quarterly 43, 1343-A1313 (2019)
43. Bhattacherjee, A.: Can Fact-Checking Influence User Beliefs About Misinformation Claims: An Examination of Contingent Effects. MIS Quarterly 47, (2023)
44. Longoni, C., Bonezzi, A., Morewedge, C.K.: Resistance to medical artificial intelligence. Journal of Consumer Research 46, 629-650 (2019)
45. Bauer, K., von Zahn, M., Hinz, O.: Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. Information Systems Research (2023)
46. Dietvorst, B.J., Simmons, J.P., Massey, C.: Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. Management Science 64, 1155-1170 (2018)
47. Riedl, R., & Stangl, F. J. : The Routledge companion to management information systems. Routledge (2025)
48. Niehorster, D.C., Andersson, R., Nystrom, M.: Titta: A toolbox for creating PsychToolbox and Psychopy experiments with Tobii eye trackers. Behav Res Methods 52, 1970-1979 (2020)
49. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K.: PsychoPy2: Experiments in behavior made easy. Behavior research methods 51, 195-203 (2019)