Mindful Machines: Enabling HCI Research Exploring Novel Brain-Driven Interaction Paradigms for Collaboration

by

Christopher Micek

A Dissertation

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Computer Science

May 2025

APPROVED:

Prof. Erin T. Solovey, Advisor
Computer Science, WPI

Prof. Yunus Telliel
Ext. Committee Member
Humanities & Arts, WPI

Prof. Lane Harrison, Committee Member
Computer Science, WPI

Prof. Craig A. Shue, Head of Department
Computer Science, WPI

Dr. Felix Putze, Ext. Committee Member
Mathematics & Computer Science
University of Bremen

Abstract

With advances in the capabilities and affordability of brain sensing technologies, brain data is increasingly being integrated as input into interactive systems. Such data can give insight into the cognitive and affective states of users, augmenting their capabilities and enriching interactions, as well as informing user-centered design and evaluation of innovative interfaces. However, there is a dearth of user-friendly tools supporting the development of brain-computer interfaces, which typically requires a high level of time and expertise. The aims of this dissertation are to develop and evaluate such tools and methods to make working with brain data accessible, and to demonstrate the utility of brain signals in novel interaction paradigms for collaboration.

Acknowledgements

This work would not have been possible without the unwavering support and mentorship of my advisor, **Dr. Erin Solovey**. Thank you for thoughtful edits, insightful conversations, and broadening my horizons to the wide, wide world of HCI. Most of all, thank you for being a constant reminder that there are, in fact, people out there who care about the actual human beings who might use the tech we make.

Thank you to my wonderful lab mates and colleagues at WPI, past and present, in the Interaction Lab and beyond—especially **Ally Howell-Munson**, **Shruti Mahajan**, **Max Chen**, and **Khulood Alkhudaidi**—for collaborating, commiserating, and conspiring in equal measure.

Thank you to my colleagues at the University of Bremen, **Dr. Felix Putze**, **Lasse Warnke**, and **Lourenço A. Rodrigues**, who were instrumental in planning and conducting our cross-continental collaboration study! I will forever be amazed we were able to collect and synchronize several streams of data from multiple people in separate locations. Thank you as well to **Henry Huang** and **Itzel Sanchez** at Harvard College for helping analyze the veritable treasure trove of data from this project.

Thank you to the faculty and my fellow graduate students in the Future of Robots in the Workplace – Research & Development (FORW-RD) NSF Research Traineeship Program, especially **Kristen Keane** for her seemingly bottomless supply of energy for organizing event programming and professional development activities, and **Prof. Yunus Telliel**, whose course and workshops helped me formalize and crystallize my passion for ethics into a sizable contribution of my research.

Thank you to the WPI Computer Science Department staff, especially **Refie Cane**, **Nan Zhang**, and **Kat Crighton**, for being the oil that keeps the wheels of the entire operation running smoothly. Impromptu stickers and chocolate made many a bad day instantly better, and obstacles in the path were never long standing when they worked their magic.

Thanks to my friends in the WPI Graduate Workers Union, for making good trouble, showing me that there are in fact people from other departments on campus, and, against long odds, building a supportive community for grad students on campus that I'm happy and excited to be a part of.

Last but never least, thank you to my family for their limitless love and support, and for believing in me even when I didn't believe in myself. Many thanks as well to my friends and roommates who keep me grounded, and remind me that there is actually a world outside the academic bubble I enjoy spending time in. I also must thank **Charlie**, our apartment cat, for her unconditional love and emotional support.

I additionally extend my appreciation to the funding sources which have supported my work. This material is based upon work partially supported by the National Science Foundation (NSF) under Grant No DGE-1922761. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Contents

1 Introduction **1**

1.1 Overview of Brain-Computer Interfaces 1

1.2 BCIs and HCI: Prior Work and Future Directions 3

1.3 Motivation 4

1.4 Dissertation Outline and Organization 4

1.5 Summary of Contributions 5

1.6 Authorship Statement 6

2 Background **8**

2.1 Accessible BCI Development 8

2.2 Multi-User BCIs and Inter-Brain Synchrony 12

3 A Model for Enhancing the Reproducibility and Reuse of HCI Research Using Brain Signals **14**

3.1 Introduction & Motivation 14

3.2 Contributions 16

3.3 Related Work 17

3.4 Literature Database Curation 19

3.5 Characterizing HCI Papers with Brain Signals 21

3.5.1 Methods 21

3.5.2 Paper Demographics 23

3.6 Experiment Model for HCI Research with Brain Signals 24

3.7 Reporting Practices in HCI Research with Brain Signals 28

3.7.1 Prevalence of Experiment Model Attributes 28

3.7.2 Structure of Experiment Reporting 29

3.8 Expert Perspectives and Review 31

3.9 Discussion & Outlook 33

3.10 Conclusion 38

4 BrainEx: An Easy-To-Use Tool for Exploring Brain Signal Datasets 40
4.1 Introduction 40
4.2 Contributions 42
4.3 Background 42
4.3.1 Functional Near-Infrared Spectroscopy (fNIRS) 42
4.3.2 Analytic Tools for fNIRS 43
4.3.3 Exploring Similar Sequences in Time Series 44
4.3.4 Efficient Sequence Similarity Search Using Multiple Warped Distances 45
4.4 BrainEx Engine Architecture 46
4.4.1 BrainEx Engine: Distributed Preprocessing Algorithm to Compute Clusters 48
4.4.2 *BrainEx* Engine: Distributed Similarity Search Algorithm 49
4.4.3 *BrainEx* Engine: Time Series Indexing and Memory Optimization . . 50
4.4.4 *BrainEx* Engine: Operations 50
4.4.5 *BrainEx* Engine: Supported Datasets 51
4.5 BrainEx Visual Exploration Design 51
4.5.1 Usage Goals 51
4.5.2 Functional Requirements 52
4.5.3 Interface Components 53
4.5.4 Similarity Search 54
4.5.5 Cluster Exploration 55
4.5.6 Feature Distribution Exploration 56
4.5.7 Integration Pipeline 58
4.6 Performance Benchmark Summary 59
4.7 Preliminary User Study 59
4.7.1 Study Design 59
4.7.2 User Study Results 61
4.8 fNIRS Case Study 64
4.8.1 Dataset Description 65
4.8.2 Case Study Results 66
4.8.3 Case Study Conclusion 67
4.9 Discussion 67
4.10 Future Work 68
4.11 Conclusion 69

5 BCIs for Teamwork: Relevant Background 70
5.1 Teams and Creativity 70
5.2 Artistic Brain-Computer Interfaces and Creativity Support Tools 73
5.3 Brain-Computer Interfaces to Support Teamwork 74

5.4 Public Perceptions of BCIs and Emerging Technologies 74

6 BCIs for Teamwork: Exploring the Relationships Between Brain Markers for Individual and Group States During Collaboration 78

6.1 Introduction 78
6.2 Study Design 80
6.2.1 Participants 80
6.2.2 Data Acquisition and Experiment Procedure 80
6.3 Analysis Methods 84
6.3.1 Task Performance 84
6.3.2 Emotion Regulation Strategies 85
6.3.3 Team Processes from Speech and Behavior 86
6.3.4 Individual EEG Metrics of Attention 89
6.3.5 Team Brain Metrics 90
6.3.6 Analyzing the Relationship Between Individual and Team Metrics . . 101
6.4 Results 104
6.4.1 Relationship Between Individual and Group Neural Measures 104
6.4.2 Influences on Team Performance 105
6.4.3 Relationships Between Individual Neural Measures and Behavior . . . 109
6.4.4 Relationships Between Group Neural Measures and Behavior 112
6.5 Discussion 119
6.5.1 Limitations and Future Work 121
6.6 Conclusion 123

7 BCIs for Teamwork: Exploring the Potential for Brain-Computer Interfaces in Collaborative Contexts 124

7.1 Introduction 124
7.2 Exploring User Needs: Design Interviews with Team Members and Leaders . 126
7.2.1 Participants 126
7.2.2 Study Design 128
7.2.3 Generative Card-Sorting with “Superpowers” to Probe Challenges Faced by Teams 129
7.2.4 Semi-Structured Interviews: What Works Well, What Doesn’t, and How Tech Could Help 131
7.2.5 Speed-Dating Possible Futures 142
7.3 Discussion and Future Work 152
7.3.1 RQ1: What unmet needs do stakeholders have during collaboration? 153
7.3.2 RQ2: In what ways could brain-computer interfaces help fulfill these needs? 153

7.3.3 RQ3: What concerns do stakeholders have about BCIs being used to support teams, and what risks do they foresee? . . . 154
7.3.4 RQ4: How could such risks and adverse impacts be minimized over the course of design and development? . . . 155
7.3.5 Future Work . . . 155
7.4 Conclusion . . . 156

8 Ethics and Broader Impacts **157**
8.1 Prior Work in BCI Ethics and Neuroethics . . . 158
8.2 Ethics in Context: Stakeholder Perspectives on the Development of BCIs for Workplace Teams . . . 160
8.3 Ethical Guidelines for the Responsible Use of BCIs for Collaboration . . . 161
8.4 A Motivating Example . . . 162
8.5 Broader Impacts . . . 163
8.5.1 Advancing Research Reproducibility and Accessibility in HCI . . . 163
8.5.2 Societal and Workplace Impacts of BCIs for Teams . . . 164
8.5.3 Community and Public Engagement . . . 164
8.5.4 Evaluating Broader Impacts . . . 165
8.6 Conclusion . . . 165

9 Concluding Remarks **166**

A Detailed Experiment Model with Examples **208**

B Team Collaboration Task Demographics **216**

List of Figures

3.1 Prevalence of attributes across all analyzed papers, values indicate the relative number of papers reporting the corresponding aspect. 29
3.2 Correlation of model attributes prevalence and paper age for selected items, sorted by correlation coefficient. 30

4.1 BrainEx is a web-based visual analytic tool, designed for exploring sequence similarity and clusters within brain signals. On the left, we see 50 similar sequences with color used to encode metadata about the search results, and on the right is a person wearing a functional near-infrared spectroscopy brain sensing cap. 41
4.2 Comparison of how DTW and point wise similarity matching occurs. DTW (on top) allows for a one-to-many mapping, as seen by points from the top sequence all mapping to a single area on the bottom sequence that occurs earlier. Pointwise (on bottom) only allows for a one-to-one mapping of points that occur at the same time in both sequences. 44
4.3 Representations and groupings derived by GDTW. The colors of the sequences correspond to clusters of similar sequences and their respective representatives. 45
4.4 *BrainEx* Pipeline Overview. When preprocessing a dataset, the time series are divided into all possible sequences of all possible lengths and then clustered into similar groups of equal length. After preprocessing, researchers can interactively explore the clusters and perform a fast similarity search by finding the cluster representatives most similar to the target sequence, and only searching the clusters they represent. 47
4.5 The BrainEx System Architecture. The first component on the left is the *BrainEx* Engine Server, developed in Python and usable with Linux OS. The middle component is the API that preprocesses datasets, performs similarity searches, and clusters data which is implemented with Django. The last component is the *BrainEx* website interface which the user accesses and is developed in HTML, CSS, React, and D3.js. 47

4.6 The Preprocess Data page is where a user will upload a new dataset and specify the parameters for preprocessing. **A)** Users must manually provide the number of header rows and feature columns; the default for each is 0. **B)** The user-defined similarity threshold is the minimum similarity requirement between sequences in the same cluster. **C)** The length of interest is the range for subsequences to be spliced; the default is 1-n where n is the maximum sequence length in the dataset. **D)** The available warped distances for similarity matching. Currently, Warped Euclidean, Warped Manhattan, and Warped Chebyshev are available, however the code is built to easily accept more warped distances. 48

4.7 Parts (a) and (b) demonstrate the difference between the naive distribution scheme and our implementation of distributive step slicing (DSS). The discs on the time series are individual data points. The curves above the data points represents the sliced subsequences. Different colored curves represent work done by different executors. Figure (a) shows the naive distribution scheme, when the number of time series is not divisible by the number of executors. This results in load imbalances. In this case, the load of the 'orange' executor is twice the amount of its fellows. Figure (b) shows a Generalized DSS for load (number of sub-sequences) balances over multiple time series; note that the executors' start index for each time series is set in a round-robin style to ensure further balancing of the loads. 49

4.8 Similarity Search. *BrainEx* Visual interface for conducting similarity search consisting of search options in the left panel (marked A) and a visualization of the dataset's sequences on the right (marked B). 54

4.9 Cluster Explorer. *BrainEx* interface for exploring the clusters consisting of a table of filtering options in the left panel (marked A), a table of the dataset's clusters and their color-coded feature distributions (marked B), and a visualization showing the overall distribution of clusters in the dataset (marked C). When a cluster is selected, the visualization changes to show that cluster's representative. 55

4.10 The three different vizualizations to explore the content of a cluster. A user will progress from the bar chart (**A**) to the heat map (**B**) then to the parallel coordinates (**C**) as they add additional features to the visualization. **A)** A bar chart displaying the distribution of a single feature in a sample set of sequences. **B)** A heat map displaying the joint distributions of two features in a sample dataset. **C)** A parallel coordinates view displaying the joint distributions between three features in a sample dataset. This visualization will be used for any number of features ≥ 3. 57

4.11 The results of the user study. For each plot, the x-axis shows possible responses and the y-axis shows the frequency of each. 61

6.1 Marks et al.'s [218] taxonomy of team processes. 88
6.2 Distribution of scores for escape room designs. Scores were grouped into five equal-width bins, ranging from the minimum score (49.5) to the largest possible score (100). The largest score received by any group was 93.5. . . . 105
6.3 Box and whisker plots for each group showing team members' prior familiarity with escape rooms. 106
6.4 Escape room design scores and prior familiarity with escape rooms for all groups. Design scores could range from 0 to 100, and are color-coded based on the minimum and maximum scores that were achieved—49.5 (the minimum) maps to red, and 93.5 (the maximum) maps to green, with intermediate values interpolated. Familiarity levels are provided for each participant (P1 or P2) at each study location (Location 1 - WPI or Location 2 - UniBremen) in the final four columns: 1 - Not at all familiar; 2 - Somewhat familiar; 3 - Moderately familiar; 4 - Very familiar. Values are likewise color-coded, with 1 mapping to red and 4 mapping to green. Blank cells indicate groups with fewer participants. 107
6.5 Box and whisker plots illustrating the relationship between design scores and the maximum prior familiarity with escape rooms for each team. 108
6.6 Counts of higher-order team process dimensions per group. 110
6.7 r^2 values from regression analyses modeling escape room design scores as a function of individual EEG indices and prior familiarity with escape rooms. . 111
6.8 Regression modeling TEI as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group. 114
6.9 Regression modeling TLI as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group. 115
6.10 Regression modeling theta band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group. 116
6.11 Regression modeling alpha band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group. 117
6.12 Regression modeling beta band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group. 117

6.13 Regression modeling MdRQA recurrence rate as a function of higher-order team process dimensions for each individual per group. 118
6.14 Regression modeling mutual information (MI) as a function of higher-order team process dimensions for each individual per group. 118

7.1 Pairwise comparison of rankings for each superpower in the second experiment session. Numbers and colors in each cell indicate the number of participants who ranked the superpower in the corresponding row greater than that in the corresponding column. Higher numbers/darker shades indicate greater agreement between participants. Note that two participants (P5 and P7) completed two different rankings, each with respect to a different team they were part of, so the maximum possible agreement is 10. Powers are sorted top to bottom in order of participant agreement on their relative importance (i.e., the sum of the respective rows). 132
7.2 Scenario 1: Align team effort and output 145
7.3 Scenario 2: Knowledge of what team members are thinking/feeling 147
7.4 Scenario 3: Detect and avoid miscommunication/misunderstanding 148
7.5 Scenario 4: Know how the team can be helped and improved 149
7.6 Scenario 5: Understand team assets and skills 150
7.7 Scenario 6: Ability to reduce stress and burnout 151
7.8 Scenario 7: Build and maintain trust/camaraderie among the team 151

List of Tables

3.1 Counts of regarded papers per year 20
3.2 Comparison of experiment representation in two papers from HCI and BCI (first pair). 25
3.3 Comparison of experiment representation in two papers from HCI and BCI (second pair). 26
3.4 Summary of the experiment model, listing all regarded attributes of HCI experiments involving brain signals. Attributes are structured in several categories. Tables A.1-A.6, in the Appendix, provide further details, including definitions and examples. 27

4.1 Participant Demographics 61
4.2 Expected and observed distributions of events from 49 clusters that have high inclusion of modeled cognitive control sequences. 66
4.3 Expected and observed distributions of channels from 49 clusters that have high inclusion of modeled cognitive control sequences. 67

6.1 Scoring rubric used for escape room designs, based on the taxonomy of digital escape rooms by Krekhov et al. [184]. A maximum of ten points could be earned for each of the first eight criteria, an a maximum of four points could be earned for each of the last four, for a total possible score of 100 points. . . 83
6.2 Emotion regulation strategies measured by the PMERQ, and the number of associated survey items. The final score for each strategy is the average of item-level responses. 86
6.3 Example prototype matrix for the driver-empath model from [117], calculated for a team with four members. (AR represents autoregression.) 93
6.4 Counts of the utterances for each higher-order dimension of Marks et al.'s taxonomy of team processes [218] and escape room design scores for each group. Utterances were summed across group members for each category; the total number of such utterances for all group members is shown in the "Total" column. 109

6.5 Pearson correlation between higher-order team process dimensions and design scores. . . . 109
6.6 TEI decision tree regression coefficients for higher-order team process dimensions. . . . 112
6.7 TLI decision tree regression coefficients for higher-order team process dimensions. . . . 113
6.8 Pearson correlation between higher-order team process dimensions and mean individual neural measures (all groups). . . . 113
6.9 Pearson correlation between higher-order team process dimensions and mean individual neural measures (outliers removed). . . . 113
6.10 Theta band decision tree regression coefficients for higher-order team process dimensions. . . . 113
6.11 Alpha band decision tree regression coefficients for higher-order team process dimensions. . . . 113
6.12 Beta band decision tree regression coefficients for higher-order team process dimensions. . . . 114
6.13 Pearson correlation between higher-order team process dimensions and mean group neural measures. . . . 116
6.14 Pearson correlation between higher-order team process dimensions and mean group neural measures (outliers removed). . . . 117

7.1 Study participants and their respective team backgrounds. Participants marked with an asterisk (*) were unavailable for the second experiment session. . . . 127
7.2 Results from the superpower card-sorting exercise from the first experiment session. Shown are the superpowers which were generated by multiple participants, and were chosen as most important for at least one. The total frequency of occurrence across all participants and the frequency of selection as the most important power are provided in the second and third columns, respectively. . . . 129

A.1 Experiment Model: Part 1. . . . 209
A.2 Experiment Model: Part 2. . . . 210
A.3 Experiment Model: Part 3. . . . 211
A.4 Experiment Model: Part 4. . . . 212
A.5 Experiment Model: Part 5. . . . 213
A.6 Experiment Model: Part 6. . . . 214
A.7 Experiment Model: Part 7. . . . 215

B.1 Team collaboration task participant demographics (Part 1). Escape room familiarity levels could range from (1 - Not at all familiar) to (4 - Very familiar), as in Section 6.2.2; the final column lists the location of each participant (Location 1 - WPI or Location 2 - UniBremen). 217
B.2 Team collaboration task participant demographics (Part 2). Escape room familiarity levels could range from (1 - Not at all familiar) to (4 - Very familiar), as in Section 6.2.2; the final column lists the location of each participant (Location 1 - WPI or Location 2 - UniBremen). Missing information from two participants who did not complete the demographic questionnaire is omitted. 218

Chapter 1

Introduction

As brain sensing technologies have become cheaper and more widespread, researchers and developers have designed systems able to leverage the unique insights into users' mental states these devices offer to create more intelligent and adaptive user experiences. Such systems have utilized users' cognitive and affective states to enhance their capabilities in several domains, including communication, entertainment, marketing, rehabilitation, and smart control of devices and interfaces [166, 367]. Applications have been developed allowing users to share their brain signals for social expression [200, 298], receive tailored recommendations in response to their valence and arousal levels [371], optimize interface layout and workflow based on cognitive load [307], and more [274, 295, 346]. In the following sections we first briefly describe brain-computer interfaces (BCIs) generally as well as existing applications and areas of study by human-computer interaction (HCI) researchers exploring brain-computer interaction, before outlining unmet needs and underexplored areas in the field that this dissertation will address.

1.1 Overview of Brain-Computer Interfaces

Broadly, BCIs fall into two main categories: those that use *explicit* or *direct* control, in which intentional changes in brain activity are used an alternative input-control mode for a digital system, and those that rely on *implicit* or *passive* input, adapting intelligently to the internal state of the user [90, 153, 310]. Although they are able to provide a host of communication and control applications for clinical populations, BCIs that use explicit control may not be as useful for healthy users, due to the typically lower bandwidth available and higher effort required compared to traditional modes of input, such as speech or text entry [381]. These systems may also be asymmetrical with respect to information exchange, similar to standard mouse-and-keyboard computer systems: computers are able to provide a wealth of information about their moment-to-moment state and capabilities, but are blind to such information about their users [127]. A computer is unable to detect when a user's condition or requirements have changed in a way that impacts the function or goals of the human-computer interaction taking place, and is thus unable to alter its behavior in real time in response these needs unless the user themselves provides a command. Passive neuroadaptive interfaces rectify this asymmetry by allowing computers to respond intelligently and autonomously to the intentions, emotions, and cognitive states of users [88], thus increasing the bandwidth available for information exchange between human and machine [91].

Brain-computer interface research in HCI builds off the concept of the biocybernetic loop [270], in which a physiological computing system collects biological signals from users, extracts meaningful information about the user (e.g., their level of engagement, heart rate, or movement speed), and responds in some way to guide them toward a particular goal state (e.g., altering the appearance of interface components to ensure engagement is above a chosen threshold, or to alter the tempo of music according based on a user's heart rate to provide a positive emotional experience during exercise) [90]. The system then continues to respond to any further changes in user state that occur in response to the initial and subsequent adaptations. However, designers and researchers have since developed several BCI interaction paradigms in addition to the closed loop described above. Along with systems that use brain signals for direct user control, BCIs have been developed that assess the mental state of the user, provide neurofeedback about the user's cognitive or emotional state, or use information about the user's brain state to evaluate a digital interface [187].

Once biosignals of interest are acquired—usually via wearable sensors connected to the user, but possibly with no-contact optical [267] or electromagnetic [329] sensing in some applications—the data are then typically preprocessed to remove artifacts and isolate signal components or frequency bands of interest. After preprocessing, the data is then analyzed to extract a control signal that is then used to trigger a response from the system, such as an event-related potential (ERP) or the mean power of a particular frequency band in the case of electroencephalography (EEG) systems enabling direct control by the user, or the detection of a cognitive or affective state of interest as determined by a machine learning model for passive systems. The response by the interface could then enable direct control by the user (e.g., the ability to control a cursor or send commands), or in the case of adaptive systems trigger either an overt event (e.g., displaying a prompt asking if the user would like help with a task) or make a covert adjustment that may be subtle or unnoticed by the user (e.g., altering the prominence of UI elements, adapting the difficulty of a video game) [90, 89]. Thus, successful development of BCIs requires the integration of expertise in several different domains, including neuroscience, cognitive science, signal processing, machine learning, and interface design, making the full development of a functional BCI challenging for non-experts.

1.2 BCIs and HCI: Prior Work and Future Directions

Human-computer interaction researchers work to improve people's relationship with technology and enhance user capabilities by designing systems that are accessible, efficient, and desirable to use. A necessary part of this design process is imagining how the current technological landscape will evolve: what kinds of technologies and interactions will be commonplace years and decades in the future? While the ability to interface directly with the human brain has been an alluring possibility in popular culture for several decades, advancements in cognitive neuroscience and brain sensing technologies have made this frontier more accessible than ever before. Faced with the prospect of more widespread adoption of this emerging technology, HCI researchers are tasked with designing novel applications which leverage its unique capabilities to enhance the lives of users of the future.

Though still in the nascent stages, work exploring the use of brain sensing has been increasingly published at HCI conferences and journals [275]. This work has examined several novel applications for brain sensing in various domains, including entertainment [283, 266, 369, 56, 377], gaming [186, 159], VR [172], education [95, 322], healthcare and rehabilitation [10, 234], robot control [192, 4], and more [262]. Some examples of novel BCI applications researchers have developed include a brain-augmented system allowing for direct and adaptive control of robots and drones [192, 4], movies that branched according to the cognitive states of observers [283, 266], and neurofeedback to help children with emotion regulation [10]. This prior work by HCI researchers has mainly focused on understanding how brain activity changes when users interact with or via digital systems, as well as developing applications using brain signals to control or moderate the behavior of such systems, but as the field has matured in the past few years, more researchers have used metrics based on brain activity (e.g., levels of workload or engagement) as objective indicators of user experiences without the need for self-report. Several studies have also explored the ethical considerations surrounding the use of BCIs and attitudes of different populations of users toward them [230, 291].

However, several areas have been neglected or underexplored in the HCI literature on neurophysiological interactions. While existing work has developed numerous novel systems using physiological data as input, there has been considerably less focus on making the design and development of these complex systems accessible for non-experts, or what best practices for research and design of these systems should be within the HCI community. Furthermore, while several interfaces have leveraged brain signals from individual users, there are few applications which are able to utilize data from multiple users simultaneously. This dissertation explores methods and frameworks for making the design of digital systems using brain signals more accessible to HCI researchers and non-experts in general, as well as forward the development of novel BCI applications to enhance day-to-day interactions and tech-mediated relationships, specifically in the context of collaboration.

1.3 Motivation

This dissertation builds on prior work in HCI and cognitive neuroscience on brain sensing and brain-computer interaction. It adapts and extends existing research and design methods, while addressing three key gaps in the current landscape:

1. The absence of **standard best practices** for HCI researchers working with brain signals, which hampers reproducibility and reuse. While HCI introduces unique research priorities that demand tailored guidance, integrating lessons from the neuroscience community can help inform the development of reporting standards for brain-computer interaction research.

2. The **high barriers to entry** for researchers and designers working with brain data, due to limited tools that make neurotechnology accessible to non-experts.

3. The **limited development of passive brain-computer interfaces for everyday, real-world applications**, especially those designed to support collaboration or make use of multi-user brain data.

These gaps motivate the two core aims of this dissertation:

1. To develop tools and methods that **support accessible, reproducible workflows for working with brain data** across the full research lifecycle—e.g., problem formation, study design, data collection, prototyping, and analysis; and

2. To explore how brain signals from individuals and groups can **enable novel interaction paradigms that enhance human relationships and collaborative experiences**, especially through ethically-designed, team-facing BCIs.

Through achieving these aims, this dissertation lays groundwork for the design and deployment of ethical, accessible, and collaborative BCI systems.

1.4 Dissertation Outline and Organization

This dissertation is organized as follows: After this high-level overview, Chapter 2 provides a review of prior work related to accessible BCI development and the use of BCIs in multi-user contexts, grounding the dissertation's aims in existing research.

Chapters 3 and 4 then each address the first aim of supporting accessible and reproducible BCI research and design. In Chapter 3, we describe our work developing a taxonomy of HCI research using brain signals to serve as a model for best practices, enabling researchers to more easily ensure their work is reproducible and reusable by others, potentially across other disciplines such as neuroscience. We surveyed a collection of 110 papers published in HCI venues which utilized brain signals to inductively construct our model to represent the breadth of current research, and used exploratory factor analysis to uncover latent patterns in reporting strategies captured by our model. We then conducted a survey of HCI and BCI experts to confirm the validity of our model, and explore potential use cases and implications. Chapter 4 describes the development of BrainEx, a GUI-driven tool allowing non-experts to explore brain signals or other time series data, and presents results of a user study illustrating that user needs for exploring and drawing insights from data were met.

Chapters 5, 6, and 7 then address the second aim of exploring the human-centered development of a multi-user BCI for enhancing collaboration. Chapter 5 begins with an overview of relevant literature regarding teamwork and creativity—our chosen collaboration context. We follow with a discussion of existing creativity support tools and BCIs for creative applications, as well as emerging work toward the development of BCIs to support teamwork, and conclude with work exploring public perceptions and ethical concerns regarding BCIs and other emerging technologies.

Chapter 6 investigates practical implementation considerations for such a support system—to design a useful and usable system, we need to ask, "What can we learn about the quality and nature of collaboration from the brain activity and behavior of team members?" To explore this question, we conducted a multi-institution study in which participants worked in teams on a realistic creative task while recording their brain activity, and report our findings. Modest links between behavioral measures of team dynamics and brain derived metrics were observed, and their potential implications and remaining future work toward developing a working support tool are discussed. Complementing this exploration of technical feasibility, Chapter 7 explores the needs and concerns of the various stakeholders that might benefit from such an intervention. Adopting a user-centered approach, we conducted a series of design interviews and brainstorming exercises with members and leaders of various teams to examine the challenges they face during collaboration, the tools and strategies they employ to address them, and possible ethical concerns with possible BCI solutions. We conclude by proposing guidelines for the ethical development of BCIs supporting collaboration, which can be used by prospective developers and researchers to ultimately design systems that provide value for teams while minimizing potential risks.

Finally, Chapter 8 explores the ethical implications and broader impacts of this work, focusing on ramifications for user autonomy, privacy, and identity, additionally providing ethical guidelines for the responsible use of BCIs in collaborative settings.

1.5 Summary of Contributions

The work in this dissertation makes strides toward improving the accessibility and reproducibility of HCI research using brain signals, and explores how this work could be actualized in the real-world context of creative collaboration. Chapter 3 provides a taxonomy of HCI research using brain signals, which can serve as a model for ensuring such work is reproducible and reusable. We have made this model publicly available and open to contributions from others, and provide guidance for its use. This chapter is adapted from work published in *Transactions on Computer-Human Interaction (TOCHI)* in 2022 [275]. Chapter 4 introduces BrainEx as an easy-to-use tool for exploring and understanding brain signal data; it is adapted from work published in the *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS)* in 2022 [139]. Lastly, Chapter 7 provides an analysis of stakeholder needs and outline design guidelines for designing ethical BCIs supporting collaboration, while Chapter 6 gives insight into the relationship between brain and behavior necessary for the practical implementation of such systems; each of these chapters will be submitted for publication in an HCI journal or conference.

1.6 Authorship Statement

The work in this dissertation is the product of years of collaboration with my fellow researchers, especially my advisor Prof. Erin Solovey. Chapter 3 is adapted from our TOCHI paper [275], and was a joint collaboration with colleagues from the University of Bremen—Suzanne Putze, Merle Sagehorn, and Felix Putze. I was not the first author of the manuscript, but I was actively involved in discussions planning and interpreting the work and contributed in several key ways: I coded all manuscripts to incorporate them into our taxonomy, refined the resultant experiment model, performed a keyword search to further categorize the reporting practices of each manuscript, wrote the section of the paper describing the emergence of a sub-field of HCI work using brain signals distinct from classical BCI research, and edited the final draft.

The work in Chapter 4 is adapted from [139] and was a close collaboration with my colleague Alicia Howell-Munson, who was the first author of the accepted manuscript. I co-wrote and edited the final manuscript along with Alicia, co-supervised the creation of the front-end GUI, and performed analysis of benchmarking data to evaluate the performance of the tool. BrainEx was also the product of work by several undergraduates who assisted on the project, including Ziheng Li and Yuncong Ma who implemented the computational backend; Michael Clements, Andrew Nolan, and Jackson Powell, who first implemented the front-end GUI as part of their major qualifying project; and James Plante, Eric Schmid, Ellery Buntel, and Yufei Lin, who built off their work; as well as Prof. Rodica Neamtu, who co-advised the project.

I am the principal author of the work in Chapters 7 and 6. The latter is a collaboration with colleagues at the University of Bremen co-advised by Dr. Felix Putze, with assistance from undergraduates at Harvard College in the United States. Lasse Warnke at Uni-Bremen is the main co-author of the manuscript. Lasse and I were both responsible for recruitment, data collection, and analysis; I focused chiefly on the analysis at the group level, and Lasse focused on the individual user data and its relationship with the other metrics. Asmus A. Eilks (Uni-Bremen) assisted with our data recording setup, and Lourenço A. Rodrigues (Uni-Bremen) assisted with study design, initial pilot testing, and interpretation of results. Max Chen (WPI), Henry Huang (Harvard College), and Itzel Sanchez (Harvard College) assisted with data annotation and analysis in the United States.

To acknowledge all my collaborators' contributions, I use first-person plural pronoun "we" throughout this dissertation.

Chapter 2

Background

2.1 Accessible BCI Development

In the past several years, researchers have made strides toward reducing the barrier to entry for designing and working with BCIs, both for amateurs just getting started as well as experts looking to meaningfully contribute to the field. Part of these efforts are in line with the emerging phenomenon of Open Science [337], a worldwide push toward making research open and accessible to others in order to enhance transparency and reproducibility and foster collaboration. Researchers have engaged in Open Science through several ways, including pre-registering their experiment and analysis methodologies (to enhance the discoverability of their work as well as hold them accountable for the conclusions they draw from the analyses they conduct), writing blog posts about their work to make it more accessible to the public, and providing open access to code and datasets. BCI researchers and the neuroscience community in particular have released several tools and datasets that others have used to advance their own work.

Among the tools that have been made available to the public to enhance accessibility to analyzing and integrating neurophysiological data are several freely available software packages, many of which are open source. EEGLAB [70] is a popular open-source MATLAB toolbox that provides users with several functions for working with EEG data across several stages of analysis, including channel and event information importing, signal preprocessing, visualization, and subsequent analysis (independent component analysis, time-frequency analysis, etc.) as well as a GUI exposing these functions in a way that is accessible to users. BioSig [338] is an open-source software library available for MATLAB and C/C++ both standalone and as an EEGLAB plugin, providing functionality for signal processing including signal segmentation and artifact detection, as well as tools for classification and import/export capabilities for a variety of data formats. Analagously, nirsLAB [365] provides a similar GUI environment and underlying functionality for working with data from functional near-infrared spectroscopy (fNIRS), which uses near-infrared light to measure the relative concentrations of oxygenated and deoxygenated hemoglobin in the brain as a proxy for neural activity. sktime [212] is an open-source Python library providing several different methods for performing time series classification, making it especially suited for developing applications which use neural data.

While these for general purpose of processing and analyzing neurophysiological data, other tools designed expressly for developing brain-computer interfaces. simBCI [199] is an open-source MATLAB framework designed to alleviate the design and testing of BCI systems by providing an environment where users can use simulated BCI data and execute signal processing and classification pipelines to more easily determine which aspects of the system are affected by the user, the equipment, and the data processing routines employed, making it easier to iterate over different possible system configurations. Similarly, SimBSI [255] is an open-source Simulink library for developing closed-loop brain signal interfaces able to leverage the Simulink interface to provide an intuitive graphical environment allowing users to perform tasks such as EEG source imaging, closed-loop neuromodulation, and cognitive task design for experiments with either humans or animals, integrating with Lab Streaming Layer for real-time signal streaming and acquisition. Finally, Turbo-Satori [208] is software toolbox enabling users to develop closed-loop BCIs and neurofeedback systems using fNIRS data, and is able to interface directly with NIRx hardware to perform signal acquisition, preprocessing, and classification in real time.

Complementing the availability of these software packages for data processing and analysis, several researchers have documented and published open access datasets of neural data differing in size, recording device, modality, and context, making them useful for benchmarking classification algorithms or testing designs for BCI systems. Among these are the Database for Emotion Analysis using Physiological Signals (DEAP) [173], which includes EEG recordings and associated peripheral physiological signals from 32 participants who rated their preferences and affective states while watching music video excerpts; the DREAMER [165] dataset of electrocardiogram (ECG) and EEG data recorded using low-cost off-the-shelf devices from 23 participants who rated their affective states while watching videos; a combined EEG and fNIRS dataset recorded from 26 participants who performed a variety of cognitive tasks (n-back, discrimination/selection response, and word generation) [302]; the Tufts fNIRS mental workload dataset [144], a large dataset with data from 68 participants who performed an n-back task; and a dataset of resting-state fNIRS recorded from 12 participants [343]. To provide a centralized platform where researchers could share such datasets, Markiewicz et al. developed OpenNeuro [217], a repository where neuroscience data from different modalities are shared in accordance with the Brain Imaging Data Structure (BIDS) community standard.

In addition to the availability of free-to-use datasets, affordable consumer-grade hardware exists researchers and enthusiasts can use to easily collect their own EEG data. The MUSE [185, 45], EMOTIV EPOC [17], and OpenBCI Ultracortex Mark IV headset with the Cyton

Biosensing Board [7] are able to record with 4, 16, and 16 channels, respectively, and are all available at price points below $1,000. Researchers have validated of each of the headsets, which have demonstrated the ability to resolve several canonical EEG characteristics, including the N200 and P300 event-related potentials (ERPs) and frontal alpha asymmetry.

Given the availability of these tools and devices, a major challenge remaining for those hoping to design useful and effective BCI systems is determining which design strategies to employ to best take advantage of them. Several researchers have investigated the quality of different BCI systems and the effectiveness of different design strategies in a clinical context to develop guidelines for developers. In general, a user-centered approach seems recommended, whereby the needs of a target population of end users are evaluated in order to inform the design of the system [196, 297, 234]. This is especially important for clinical populations, where BCI paradigms that work well for typical users or some patient populations may not be as effective for others, e.g., in a case study by Schreuder et al. [297], where patient with a brainstem ischemic stroke was able to use a BCI controlled via visually evoked ERPs, but not auditorily evoked ERPs, for text entry and item selection. Similarly, Lightbody et al. [196] employed a user-centered approach to design a BCI for use in a domestic environment by those with physical disabilities or brain injuries, engaging with potential users and healthy controls directly to help design and evaluate the system. The researchers were able to utilize feedback from users and their caregivers to identify pain points in their interface design (e.g., need for an EEG setup that was easier for caregivers), as well as features desired in future iterations (e.g., control of multimedia and smart devices). This allowed researchers to ultimately design a system that was usable and useful for the target population, rather than merely a proof of concept. Such a design process is in line with recommendations from rehabilitation professionals attending a workshop on the design of assistive BCI systems [247], who advocated for including users in the BCI design process, and provided guidelines for developing assistive systems that encourage designers to consider the practical and human problems users face in daily life rather than concentrating solely on technological problems. More generally, Mason & Birch [220] propose a functional model for BCI system design, consisting of the user, operating environment, and system components, along with associated feedback paths. Though developed with clinical systems in mind, this work provides a common taxonomy that other non-clinical systems can build from.

Beyond a clinical context, researchers in human-computer interaction and other disciplines have compiled recommendations and reviews for tools, and created systems and frameworks to demonstrate novel interaction paradigms and enable development by others. Such work includes reviews of the different classification algorithms used in a BCI context (with a focus on considerations especially relevant for BCI applications, such as online classification and transfer learning) [204] and different software tools available to assist with BCI development, including those for developing web-based BCI systems [312], as well as explorations of how brain signals have been integrated in particular contexts, such as user experience (UX) and neuromarketing research [195] or virtual reality (VR) games and experiences [205]. HCI researchers have also expanded on prior work with their knowledge of interface design methods and user research to model user interactions with BCI systems [280] and conceptualize novel reinterpretations of existing interaction paradigms. Williamson et al. [360] developed Hex-O-Spell, an iteration upon a traditional BCI speller designed to increase the information bandwidth of users, by arranging character groupings in a radial menu with selection controlled via motor imagery. The design has since been adapted and enhanced by others, e.g., by modifying target selection to use ERPs [331], and combining it with rapid serial visual presentation (RSVP) of images to build a BCI-based framework for image search on the web [190]. Semertzidis, Zambetta, & Mueller [299] provides a critique of the traditional "command-response" design of BCI systems, and envisions a future which goes beyond mere"interaction" toward human-computer integration, in which humans and digital technology work together toward the same or complementary goals. The authors explored the design space of this paradigm by developing three different prototypes as case studies: Inter-Dream, an interface to guide participants toward healthy sleep states; Neo-Noumena, a system to communicate participant emotional states to users and others via mixed-reality displays; and PsiNet, a system designed to amplify inter-brain synchrony across users using transcranial electrical stimulation.

Based on their design prototypes, Semertzidis and colleagues [299] developed a design framework for their brain-computer integration paradigm, envisioning systems as existing on a two-dimensional plane, with "neural congruence" (the degree to which the experience the BCI is encoding from brain activity is similar to that the user experiences when decoded by the system) on one axis and "distribution of agency" between multiple users, or between the user(s) and the system, on the other. Other HCI researchers have also developed similar frameworks and taxonomies. Vasiljevic & de Moranda [334] developed a model, taxonomy, and methodology recommendations for research exploring EEG-based BCI games to aggregate the terminology and methodologies used in the BCI and game domains. Kosmyna & Lécuyer [176] developed design guidelines for BCI guiding systems (i.e., how users are instructed on how to use BCI systems), examining the *feedback* and *feedforward* aspects of these systems and their temporal, content, and medium characteristics. More generally, they also developed a multidimensional feature space to help characterize, compare and design EEG-based BCI systems [177], largely enumerating possible design choices along the dimensions of signal acquisition, signal processing, classification, and the type of BCI application. Jeunet et al. [157] likewise developed guidelines for researchers conducting experiments with BCIs enumerating possible design considerations in several relevant dimensions (signal acquisition, data processing, and experiment design). However, while this work is undoubtedly useful for researchers and developers looking to design useful BCIs, they tend to focus on BCIs used in specific domains (e.g., games), or else examine BCIs more generally and not in the context of HCI research. The guidelines that do draw from insight into both the BCI and HCI communities [176, 178] are geared more toward placing work in a particular design space, rather than ensuring that research is reproducible by others in either community. To develop a model that does so, a systematic accounting of the existing literature in the HCI community using brain signals is necessary.

Along with design guidelines, HCI researchers have also developed software and hardware tools to educate novice, non-technical users about brain-computer interfaces and physiological sensing. These include environments for students to design and experiment with brain-controlled applications [61, 126], tangible interfaces allowing users and others to visualize their brain activity [100], low-cost brain sensing hardware [14, 26], and crafting toolkits making it easy for users to create interfaces built from thermochromic fabrics responding to their physiological arousal [332]. While useful for introducing the relevant technical concepts to novices and giving them the freedom to implement their ideas for interfaces, their capabilities may be limited, and do not allow users to deeply explore and manipulate their data, or iterate over multiple experiment and interface designs quickly. However, several tools from the HCI and neural engineering communities enable rapid prototyping of BCIs and other interfaces using physiological sensing. Several of these tools are programming environments pre-packaged with analysis methods useful in multiple stages of the BCI lifecycle (e.g., signal acquisition, preprocessing, feature extraction, and classification) as well as the ability to easily test and evaluate new methods[294, 222], with some capable of performing online analysis for real-time systems [118, 181], integrating with VR headsets and virtual environments [285], and even allowing collaboration and sharing data with other researchers on the web [211]. However, despite these capabilities, a modular framework for rapidly prototyping BCIs able to interface *with* websites or web-based applications does not currently exist.

2.2 Multi-User BCIs and Inter-Brain Synchrony

Somewhat separate from the question of what best practices and tools are available for designing BCI systems is what useful and innovative applications do they enable, and in which contexts are they well-suited. Collaboration, in particular, has emerged as a promising area for further exploration, with insights into users' mental states enabling the creation of more personalized and supportive work environments. As many day-to-day interactions, whether mediated by technology or otherwise, are social in nature, a major area of interest relevant to members of the HCI community employing brain signals to drive novel interactions is work from *social neuroscience* — an interdisciplinary field studying the underlying biological mechanisms of social interactions and cognition. Though historically relying almost exclusively on experiments with single participants, studies which employ hyperscanning, the recording of brain or other physiological activity simultaneously from multiple participants, has become more prevalent in recent years [50]. Such studies typically rely on a measurement of synchrony between the brain activity of users to explore the dynamics of different types of social interactions. This interpersonal synchrony is typically task-independent, and has been observed in several different types of tasks with a social component [64], including team Jenga [201], tangram puzzles [245, 246], solving math problems, finger tapping, singing, drawing [64], word games [249], group brainstorming and creative problem solving [206, 366], and more. Although greater synchrony has tended to correlate with greater team performance, this is not universally the case [49]; Dommer et al. [77] found that participants experienced greater inter-brain synchrony when working together in a joint n-back task versus individually, despite achieving worse performance, indicating that the social nature of the task plays a role in this phenomenon [64, 49]. Teams are complex metastable dynamical systems, and oscillating between states of high and low synchrony is typical in realistic interactions [224]. The studies above measure inter-brain synchrony by measuring the *coherence* [113] between participants' brain activity, but other techniques exist as well; Eloy and colleagues [85] demonstrated that fNIRS can be used to measure levels of team processes important for collaboration in real-time by employing multidimensional recurrence quantification analysis (MdRQA), a technique able to identify patterns in multidimensional data without a priori assumptions, making it an ideal technique to employ for quantifying team processes across a variety of contexts.

Despite the numerous studies examining brain synchrony in different social contexts, very little work has been done employing brain sensing specifically in the context of group creative collaboration, and even less has employed hyperscanning to leverage data from multiple users simultaneously (one example is Morgan, Gunes, & Bryan-Kinns [235], who measured the behavior, heart rate, and brain activity of pairs of drummers during collaborative music making). Of the work that exists, the majority of BCIs leveraging inter-brain synchrony have been for art pieces or proof-of-concept systems exploring social connection. PsiNet, developed by Semertzidis, Zambetta, & Mueller [299], was designed to explore the interpersonal integration of consciousness through brain-to-brain integration of participants. Participants' brain activity was recorded, and synchronized between group members using transcranial electrical stimulation, allowing users to experience others' brain signals as if they were generated by their own body. The Magic of Mutual Gaze was an art installation and performance piece for two participants seated across from each other. EEG headsets are used to record the participants' brain waves, and visuals are displayed capturing the spontaneous synchronization of their brain activity [272]. Similarly, BrainWave Drawings by Nina Sobell displays a real-time portrait of two participants who brain waves are recorded via EEG, which becomes distorted when their brain waves diverge [272]. Though such art pieces have provided thought-provoking opportunities for examining the physiological underpinnings of social connection, the practical applications for a BCI to assist teams with collaboration remains largely unexplored.

Chapter 3

A Model for Enhancing the Reproducibility and Reuse of HCI Research Using Brain Signals

Abstract

In HCI, there has been a push towards open science, but to date, this has not happened consistently for HCI research utilizing brain signals due to unclear guidelines to support reuse and reproduction. To understand existing practices in the field, this chapter examines 110 publications, exploring domains, applications, modalities, mental states and processes, and more. This analysis reveals variance in how authors report experiments, which creates challenges to understand, reproduce, and build on that research. It then describes an overarching experiment model that provides a formal structure for reporting HCI research with brain signals, including definitions, terminology, categories, and examples for each aspect. Multiple distinct reporting styles were identified through factor analysis and tied to different types of research. The paper concludes with recommendations and discusses future challenges. This creates actionable items from the abstract model and empirical observations to make HCI research with brain signals more reproducible and reusable.

3.1 Introduction & Motivation

In human-computer interaction (HCI), there has been growing interest in integrating brain sensing into interactive systems. The field is developing traction and we anticipate further expansion as sensor technology becomes more practical for interactive applications. While HCI researchers have demonstrated many novel ways of integrating brain data into HCI practice, there are several challenges that slow progress toward reaching the potential for real-world use.

In particular, HCI research with brain signals requires a diverse skillset, including understanding of brain function, signal acquisition methods, signal processing, and machine learning, in addition to HCI, design, and software engineering. Each of these aspects are themselves active research areas. This steep learning curve and interdependence of numerous research areas limits the pace of innovation. To overcome this challenge, researchers often attempt to build on each other's work. However, this has been difficult to date, and it is more common to see each research group creating their own datasets, analysis pipelines, and following their own methodologies to move their research forward.

A contributing factor to the limited research reuse is the differing norms stemming from research cultures in related fields, making it challenging to publish, reuse, and reproduce work in HCI venues. Reviewers often have contradictory expectations, depending on their background. In addition, page limitations and cultural norms result in valuable details omitted that could support reproduction, reuse, and extension of published HCI research.

As an example, we consider the creation of adaptive interactive systems that modify system behavior based on a cognitive or affective state, measured by brain sensing. Developing such systems requires real-time processing pipelines for brain input, including infrastructure to capture, transmit, and interpret neural signals. It also involves designing appropriate experimental paradigms for data collection, implementing and tuning user interface components, devising adaptation strategies for responding to brain data, integrating all components into a cohesive application, and conducting user evaluations to assess performance and usability.

Currently, researchers start from scratch when implementing and studying such systems. This is not only very costly and time-consuming, it also creates opportunities for errors and requires extensive expertise in a number of different fields. Under such conditions, reproducible research and sharing of resources and datasets are especially important for effective progression of the field. Once research is under review in HCI venues, some reviewers may expect rigorous experimental controls as in neuroscience, while other reviewers may expect clear demonstration of usability, and still others may value highly novel and visionary proof-of-concept demonstrations over rigorous experiments. In HCI, contributions can be made to the field from many of these perspectives, and more, but it can be challenging to understand what is valued.

To move towards better reproducibility and sharing of resources, there are two related gaps that need to be filled. First, there is a need for systematic exploration of the approaches, methods, and applications in HCI research with brain signals today, to identify common themes and potential for re-use. This involves identifying recurring domains and applications, types of modeled cognitive and neurological states, methods of integrating brain signals in HCI research, and the method and sensor for measuring brain activity. Secondly, it is important to study the current practices for presenting and documenting work in this research area to ensure the feasibility of reproduction and reuse. Both of these goals require the collection and curation of a large database of HCI publications that specifically focus on the use of brain signals in HCI research.

3.2 Contributions

This chapter is an abridged version of our TOCHI manuscript [275] focusing on the development of our experiment model for HCI research using brain signals and its implications for reproducibility and reuse. For further details about specific examples and analyses, we refer the reader to the full paper. Here we report on the following:

1. The curation of a dataset of 110 HCI papers published since 1996 that explore brain sensing. Through analysis of these published works, we can look at the current state of HCI research and reporting practices when working with brain signals.

2. Systematic analysis of the dataset to compare domains, applications, modalities, measured mental states and processes. *This work substantiates existing manifestations of HCI research with brain sensing today and identifies commonalities as well as heterogeneity within this emerging, independent research area.*

3. Coming from the broad variety of research identified in (2), we created an overarching experiment model that provides a formal structure for describing HCI research with brain signals. The experiment model serves as a starting point for further analysis on how experiments in this field are reported and what aspects may be of relevance (See (4)). The model provides definitions as well as examples for each aspect and is evaluated by a group of external experts. *This step creates an abstract foundation of experiment reporting beyond individual papers.*

4. We compare the experiment model with the reality of the curated dataset of HCI papers. We perform statistical analysis to investigate the structure of experiments and experiment reporting in a data-driven way. *This provides evaluation of the model as well as analysis of reporting structure and reporting gaps, and reveals connections to the different sub-groups identified in (2).*

5. Informed by our data-driven analysis of current practices within and outside of HCI, we derive a set of recommendations and considerations for future authors and reviewers and discuss future challenges. *This creates actionable items from the abstract model and the empirical observations in the analysis.*

3.3 Related Work

Multiple processes have been developed that support the goal of reproducible research, such as citable open data repositories and pre-registration (such as the Open Science Framework[1] or Zenodo[2]). More and more, publication outlets and funding agencies expect the publication of recorded data to ensure that other researchers can check the reported results, but also benefit from the investment of time and money. It should be noted that open data (in the sense of releasing purely the raw recorded data) on its own is not enough to ensure replicable research: If the data is not well-documented, the ability of independent researchers to reproduce or build on it is restricted. This challenge in the curation process of open data is related to the *long tail of science*. This term describes the phenomenon that most data is produced in small, individual research projects with specific applications, domains, or restrictions in mind (see [93] for an analysis of this phenomenon in the field of neuroscience).

Making this data accessible to other researchers requires a joint understanding, or even better, *a formal model* of how the data should be shared and what information is necessary to use the data successfully. Such an experiment model provides a list of mandatory or optional attributes, each of which describes a specific aspect of an experiment. Attributes are structured in categories of related items. Often, these models are associated with software ecosystems that support the documentation, publication, and querying of data. As even minimal models with few attributes were shown to prevent common documentation gaps in the literature [227, 123], over 50 publishers, e.g., *Nature* and *PLoS*, already recommend using such a model as part of the publication workflow[3]. Disciplines such as genomics [160], systems biology [362], enzymology [327], and materials science [279] have all developed such models which allow the description of experiments in a common language on the basis of common attributes, enabling re-use of data.

In other fields, some of which are closely related, there are standard practices in place to facilitate reproduction and reuse of prior work, building a solid foundation that later research can build on. For example, the field of neuroscience has invested much effort into the creation of standards for published data sets with a joint documentation model, such as the OpenNeuro initiative [4]. Assessment of the impact of shared brain imaging data on the scientific literature has found that sharing data can "increase the scale of scientific studies conducted by data contributors, and can recruit scientists from a broader range of disciplines" [233]. Babaei et al. [16] extracted published practices of EDA recording and analysis and published a structured and modifiable version of the resulting model[5] and similarly, Bergström et al. [24] analyzed reporting strategies for Virtual Reality experiments and

[1] `https://osf.io/`
[2] `https://zenodo.org/`
[3] `https://www.beilstein-institut.de/en/projects/strenda/journals/`
[4] `https://openneuro.org/`
[5] `https://edaguidelines.github.io/`

distilled a checklist for future publications. For structured publication and documentation of experimental processes, approaches like open electronic lab notebooks (e.g. "Open Lab Book"[6]) have emerged, offering this opportunity. In the field of machine learning, which is often applied to brain signals, it has become increasingly common to publish executable code on the basis of open toolkits and programming languages, often on version control platforms such as GitHub. This overcomes the difficulty of reporting all algorithm parameters in a page-limited paper. The "Papers with Code" database [7] provides a repository for publications with associated code.

There also exist unifying initiatives in the brain-computer interfaces (BCI) community, which has some overlapping as well as some clearly distinct goals from the HCI community. The BNCI Horizon 2020 initiative [37] is a large-scale effort for standardization in conceptualizing and communicating of BCI technology. A difference to the approach in our paper is that the BNCI initiative is mostly driven by the BCI community, and focuses strongly on aspects of signal processing and machine learning. We will see later that many applications of brain signals in HCI are not BCI-related (i.e. do not involve real-time processing of brain activity) and we also argue that the HCI community of researchers using brain signals overlaps with the BCI community to some extent, but is ultimately different from it and exhibits different needs for publication. The BNCI initiative outlines four main challenges, including the lack of a "common terminology" and the absence of "curated benchmark data sets." It also acknowledges that "the BCI community might not be fully aware of relevant work in other fields," highlighting the need for communication in interdisciplinary formats, which are a staple of the HCI community. In one of the few conceptual works in the field, Kosmyna and Lécuyer [175] establish a conceptual framework to characterize different works using brain signals along four axes (some with sub-axes) related, for example, to the type of input or the underlying neural mechanism. Many other reviews in BCI literature summarize parts of the research landscape systematically: For example, Aricò et al. [12] discussed passive BCIs that are developed for use outside the lab, and Lotte et al. [205] explored BCI in combination with VR technology. While these reviews give a good overview on the state-of-the-art methods, they do not have a particular HCI focus and do not discuss experiment conduction and reporting. In contrast, a recent monograph provides an introduction and review of brain input research from an HCI perspective [310], but is not centered on experiment reporting, reproducibility and reuse. In comparison to past publications, our paper focuses on a tailored subgroup of papers, namely HCI-centric work which uses brain data. This focus gives us the opportunity to dive deeper into the specific characteristics of these papers and their reporting practices. However, this does not mean that the other meta analyses in BCI are not applicable; indeed, we find that many of their observations regarding good research practices are a common thread which we also encounter here.

[6] `https://openlabnotebooks.org/`

[7] `https://paperswithcode.com/`

The Association for Computing Machinery (ACM) has introduced an alternative approach to strict enforcement of particular practices. It instead uses positive reward mechanisms through the concept of badges that reward the publication of artifacts, such as code or data, for reproducing research results[8]. This approach has seen adaption at some ACM venues, such as ACM RecSys[9], however it is not yet widespread at HCI conferences. We think that is because HCI is not a purely computational discipline, but a field that combines computational, experimental, analytical, and human-centered aspects; thus, it may not be clear what information is needed and expected for reproducible research.

Still, efforts toward reproducible research and an open data culture have reached the HCI community as well and have sparked discussions about the utility and necessity of replication studies in HCI [136][10], the benefits and challenges of open source and open data publishing [81], the use of pre-registration [58]. This resulted in changes to incorporate reproducibility and transparency in the review process of HCI conferences, such as ACM CHI[11]. Besides CHI and ACM, several other publishers which are relevant to the HCI community[12] encourage their authors to publish their experiment protocols and refer to them within the corresponding manuscript. Despite such initiatives on an organizational level, Wacharamanotham et al. [348] investigate the status quo in transparency of CHI research artifacts and come to the conclusion that most researchers do not publish their research artifacts due to misunderstandings about possible options and the reasons behind making them available. Echtler and Häußler come to a similar conclusion with regards to open source publication of HCI software [81]. We find that in a diverse and interdisciplinary field such as HCI, it is not always clear what this should look like and how it applies to each sub-field within HCI. By examining current practice in the HCI subfield using brain signals, our paper takes initial steps toward building this understanding.

3.4 Literature Database Curation

In this section, we describe our methodological approach to collecting and curating a large database of publications on the use of brain signals in HCI research. We created the database with two central goals in mind. First, we wanted to identify how the field is structured in terms of domains, employed sensor technology, common brain input paradigms, recurring cognitive and neural constructs, and other attributes. Additionally, we looked at factors which are often tied to reproducibility, such as the number of participants and the number of times the data or artifacts of a paper were actually building on another study. The methods

[8] `https://www.acm.org/publications/policies/artifact-review-badging`
[9] `https://recsys.acm.org/recsys20/call/#content-tab-1-1-tab`
[10] `http://www.replichi.com/`
[11] see `https://chi2020.acm.org/blog/changes-to-the-technical-program/#more-1870`
[12] E.g.,`https://plos.org/open-science/`, `https://www.hindawi.com/journals/ahci/guidelines/`

Table 3.1: Counts of regarded papers per year

Year	Number of Papers	Year	Number of Papers
1996	1	2011	9
2001	1	2012	8
2002	1	2013	7
2003	1	2014	13
2004	3	2015	16
2006	1	2016	11
2007	1	2017	10
2008	2	2018	8
2009	6	2019	8
2010	3		

and results of this investigation will be presented in Section 3.5. Second, we wanted to systematically categorize and statistically analyze the current reporting practices, looking for commonalities as well as gaps, with an eye toward reproducibility and reuse of brain signal research in HCI. The methods and results of this investigation will be presented in Section 3.7.

We considered publications that made use of brain signals in combination with HCI methods. This often falls into two categories: 1) the application of HCI methods to systems employing brain input (e.g. usability evaluation of a brain-based adaptive system), or 2) the use of brain signals to create or enhance HCI methods (e.g. using brain signals to evaluate workload induced by a user interface). In the selection, we aimed to exclude publications which focus solely on the processing of brain signal data to study signal processing or machine learning techniques, without an intention to employ the results to HCI-related research questions. To identify such publications, we included only peer-reviewed papers that were published at dedicated HCI conferences or journals; see the full manuscript for the complete search criteria and list of venues [275].

After applying our search criteria, we ended up with 110 papers eligible for the survey. The created database is available as supplementary material to the paper.

We are aware that our restriction to HCI-focused venues excludes some potentially relevant papers from the analysis, for example ones focusing on BCI itself, neuroergonomics, or specific domains to which brain signals can be relevant. While such works can be very important from an HCI perspective, identifying them among other, less HCI-oriented papers in said fields, is very difficult without formal filtering criteria. We made this decision to ensure that the authors themselves considered HCI to be the central contribution in their work, thus minimizing the chance of adding "false positives" to the database. By publishing research at an HCI venue, authors signal that they consider their work to be most impactful in the field of HCI. They also use this decision to select a specific target audience, namely the HCI community for which brain input is one tool among many, in comparison to a venue dedicated to such systems. As these conferences and journals employ HCI experts as reviewers and editors, we can assume that the accepted manuscripts indeed have a focus on HCI. Additionally, the selected venues all put an explicit focus on interdisciplinary research, which makes it more likely that the publications address a wider audience. The fact that other HCI-oriented research on the use of brain data exists does not limit the generalizability of our results as we still cover a sample of 110 papers with an in-depth analysis.

3.5 Characterizing HCI Papers with Brain Signals

To gain a better understanding of what HCI research contributions with brain signals look like today, we systematically explored the curated literature database. We were interested in characterizing the diversity and common themes across published HCI papers, the relationship to other fields, and the potential for reproducibility and re-use.

3.5.1 Methods

For each paper, we collected a number of "paper demographics" to characterize the state of HCI research with brain signals. On the one hand, we were interested in aspects which provide a general context of the paper to study different types of investigations. On the other hand, we looked at aspects related to reproducibility of the reported research. These demographics include:

- *Application/Domain*: For what domain or specific application is the brain signal used? (e.g. education, robotics)
- *Cognitive/affective state or process*: Measured brain activity is usually inspected for specific patterns related to specific cognitive or affective states or processes. (e.g. error potentials, cognitive workload)
- *Brain signal integration type*: These are described in more detail at the end of this section, and include explicit control, implicit closed loop, implicit open loop, neurofeedback, mental state assessment, and HCI evaluation.
- *Brain signal modality*: Which type of signal is used to capture brain activity? (e.g. EEG, fNIRS)
- *Main contribution*: From the abstract, we extracted the sentence that summarizes the main contribution of the paper.
- *Number of participants*: We included this item because a frequent point of discussion in reviews and studies on reproducibility is the number of participants in an experiment and the resulting power of the analysis.

- *Is follow-up*: Is this paper a conceptual or concrete follow-up to an earlier publication of the same or other authors?
- *Exploratory or Confirmatory*: We look for evidence for whether the analysis is *confirmatory*, i.e., driven by explicit *a priori* hypotheses, or *exploratory* analysis to generate new hypotheses.

The first four categories were inspired by an analysis of the 218 unique keywords authors assigned to their papers. Apart from a small number of specific methodological keywords referring to the type of analysis employed in the paper, all keywords could be grouped into these categories. It is interesting to note that only four generic keywords "brain-computer interface" (42 times), "EEG" (49), "fNIRS" (17), and "human-computer interaction" (13) occur more than 10 times, showing the large variety of topics covered by these papers.

The employed definitions for the types of *brain signal integration* are based on the characterization presented by Krol et al. [187], extended by additional categories which play an important role in HCI (*neurofeedback* and *evaluation*)[13]. The categories can first be differentiated in *online* and *offline* use cases of brain input. The online ones are then further differentiated by how the measured brain activity is exploited. In particular, some systems use *explicit* or *direct* control and others use *implicit* or *passive* input [308, 380, 379, 63]. With *implicit* input, brain signals that occur naturally are detected passively in real-time, with no special effort from the user. Unlike most *explicit* control systems, which often utilize brain data as the primary system input, *implicit* input paradigms frequently integrate brain signals as a secondary input channel to interactive systems. Implicit paradigms can further be broken down into *open loop* and *closed loop* systems. Thus, the categories that we used for integration type are described below.

- **Explicit Control:** Also referred to as *direct control*, paradigms in which mental commands (such as motor imagery, directing the attention towards a blinking pattern) are given by the user and these are mapped directly to user interface operations (e.g. cursor movement or letter selection). These are sometimes categorized into *active control* and *reactive control* paradigms [382], depending on whether brain activity is controlled independent of any external stimuli or in response to external stimuli, respectively.
- **Implicit Closed Loop:** The initiated response to specific aspects of the measured brain activity is designed to directly influence these aspects through *online* system behavior changes. For example, a closed loop system could measure workload and adapt the interface specifically to adjust the user workload.

[13] [187] also defines the concept of "Automated Adaptation". We did not encounter any paper that qualifies for this kind of system during our analysis.

- **Implicit Open Loop:** *Online* use of brain input in which specific cognitive or affective states or processes from the data is detected and used to inform and adapt an interactive system. Open loop refers to the absence of any direct or intended coupling of the adaptation back to the input. For example, when detecting changes in workload, an open loop system may notify the user.
- **Neurofeedback:** *Online* use of brain signals which visualizes or otherwise backchannels aspects of the captured neural activity (in raw or processed form) to the user, enabling them to self-regulate their own brain activity consciously (in contrast to processing the brain activity to adjust the interface as for implicit closed loop).
- **Mental State Assessment:** *Offline* processing of brain sensor data for the classification of specific cognitive or affective states or processes from the data. This analysis is performed as a self-contained methodological contribution, without leveraging the result further. The goal is often to transfer the results to an online system at a later point.
- **HCI Evaluation:** *Offline* processing of brain sensor data with the explicit purpose of evaluating stimuli or a (non-BCI) HCI system offline from brain signal data. While the previous types of brain signal integration describe HCI approaches that employ a brain input component, evaluation is concerned with the use of brain signals to analyze general HCI systems. In contrast to mental state assessment, this is not done with the intention of explicit integration in the system control loop and usually focuses on low-level neural responses in contrast to high-level cognitive states.

3.5.2 Paper Demographics

Our goal in this section is not to report empirical findings per se, but to understand the breadth of use cases, constructs, modalities, and methodological patterns occurring in HCI research using brain signals in order to inform the experiment model we propose in the following section. We summarize the key findings of this analysis below, but refer the reader to the full manuscript for a more in-depth discussion.

The earliest paper that we identified is from 1996 [335], outlining a vision of using brain signals among other methods as a control signal for human-computer interfaces. It took five years for the next paper [76] to appear and five more for the third one [193]. Since then, the number of papers related to brain signals in HCI increased steadily over the years.

The surveyed literature varies widely in both application domains and the mental states modeled using brain data. Despite this diversity, most papers focus on constructs like attention and workload, often using EEG (particularly low-cost consumer-grade devices), with limited use of fNIRS or multimodal combinations. There is no single dominant application area, although entertainment, education, and gaming are among the most commonly explored.

The type of brain signal integration employed likewise varied considerably. While several papers detailed systems that were responsive to user brain signals, either via explicit control or open- or closed-loop implicit adaptation, most systems did not achieve full integration with real-time adaptation, often due to technical and methodological challenges. The majority of studies used brain signals for mental state assessment or HCI evaluation, typically in offline settings.

In comparing papers from HCI and traditional BCI communities, we found notable differences in motivation, methodology, and evaluation. HCI papers tend to emphasize user experience and system-level outcomes, while BCI papers focus more on classifier performance and low-level signal fidelity. Very few HCI papers offered replication, reuse, or follow-up on previous work, and data/code sharing remains rare. Tables 3.2 and 3.3 provide explicit examples for comparison.

Finally, although exploratory approaches dominate the literature, there is little standardization around confirmatory practices such as hypothesis pre-registration, power analysis, or formal effect size reporting. This further reinforces the need for clearer reporting standards and reproducibility guidelines tailored to HCI contexts.

3.6 Experiment Model for HCI Research with Brain Signals

In examining papers covering a broad variety of use cases, it becomes apparent that different publications contain different approaches of reporting the experiments, even within the field of HCI. One aspect may be covered in great detail in one paper and only briefly touched upon in another. Sometimes, this is due to the type of research contribution (e.g., whether it involves the application of machine learning techniques), while other times it is due to cultural differences (e.g., method-driven vs. application-driven research).

Thus, to investigate how HCI experiments involving brain signals are described in the literature, we first define a *model* of such experiments. The benefit of a model is that it provides vocabulary and structure for comparing different works and reporting styles and can act as a guideline on what other researchers report and expect to be reported. An experiment model is a superset of items, or *attributes*, occurring multiple times in the relevant publications, which is grouped into different categories. Each attribute defines a unit of information related to the experiment. Not all attributes are necessarily applicable to every paper. However, they should be general enough to appear regularly and across a number of applications and use cases.

The model we discuss here was created in an iterative process during initial passes of the surveyed literature: A superset of reported attributes was created, similar concepts (or identical concepts with different names) were grouped together and categories were defined

Table 3.2: Comparison of experiment representation in two papers from HCI and BCI (first pair).

Hci1 [321]	Bci1 [238]
Motivation	
Targeted "appropriate, effective responses" in "educational settings"	Expressed the goal to "facilitate greater accuracy" of the (active BCI) systems they want to improve, i.e., they aim for a more technical in a more focused area of application.
Setting	
Implemented their model in a humanoid robot that performs real-time adaptation; participant interaction is structured coarsely.	Collected data for offline processing in a design structured into a sequence of short, repetitive trials. Participants performed various mental tasks with different levels of difficulty.
Both studies were conducted in a laboratory setting.	
Construct Differentiation	
Took a relatively broad strokes approach for the definition of attention, explicitly stating that the terms engagement and attention are used "interchangeably".	Differentiates three related, but distinct mental states (fatigue, frustration, and attention) and manipulates them individually.
Signal Processing	
Described a custom mapping algorithm that builds on the "proprietary" preprocessing of the consumer-grade EEG hardware.	Elaborated the parameter values for how their signals were filtered and digitized, but goes on to state that "all aspects of data collection and experimental procedure were controlled" by the BCI system that was used. Choices of electrode placement, reference, and signal decimation were experiment parameters that were manipulated over the course of data collection.
Metrics of Success	
Looked at objective and subjective outcome measures, such as task success rate and attitude towards the agent or motivation.	Assessed and compared classification performance using the accuracy score.
Validation Approaches	
Used a manipulation check to test whether the attention model reacts to different situations. They further look at the participants' objective and subjective responses to the robot behavior and differentiate between genders.	Discussed the neural mechanisms potentially underlying the observed classification performance; the authors also discuss how electrode selection and electrode reduction influence the classification performance.

Table 3.3: Comparison of experiment representation in two papers from HCI and BCI (second pair).

Hci2 [376]	Bci2 [188]
Motivation	
Developed a “multi-touch table using [a] P300-based BCI” to explore “the embedding of BCI in new HCI situations.”	Sought to “improve classification performance” of a traditional P300 speller.
Setting	
Participants performed analogous selection tasks using both a traditional P300 speller and a new “multi-touch system” that relied on a multi-touch table to detect the presence of objects and flash the areas beneath them.	Required participants to perform a selection task using a traditional P300 speller.
Both studies used within-subjects, trial-based experiment designs, and performed both offline and online classification.	
Construct Differentiation	
Does not greatly differ across the two papers; both describe how participants must “attend” to a target stimulus in order to select it with their respective P300 systems, but do not attempt to specifically define or measure “attention.”	
Signal Processing	
Used MATLAB software provided by the BCI headset vendor to handle signal acquisition and classification.	Elaborated the parameter values for how their signals were filtered and digitized, but goes on to state that “all aspects of data collection and experimental procedure were controlled” by the BCI system that was used. Choices of electrode placement, reference, and signal decimation were experiment parameters that were manipulated over the course of data collection.
Metrics of Success	
Sought to show that participants who used their multi-touch P300 system selection success that was comparable to a traditional P300 speller.	Aimed to determine the set of recording and classification parameters that resulted in optimal classification accuracy for a standard P300 speller paradigm.
Validation Approaches	
Compared the online classification performance of their multi-touch P300 system to a P300 speller, as well as to speller performance from a prior recent study.	Analyzed offline classifier accuracy in a factorial fashion across multiple recording and feature configurations to determine the optimal set of parameters, then validated their choice with another, online classification experiment.

Table 3.4: Summary of the experiment model, listing all regarded attributes of HCI experiments involving brain signals. Attributes are structured in several categories. Tables A.1-A.6, in the Appendix, provide further details, including definitions and examples.

Category	Experiment Attributes
Technical aspects of recording	type of sensor, sensor position, sampling rate, measurement quality, reference, auxiliary signals, synchronization with stimuli and other signals, recording environment
Task description	participant restraints, output devices, input devices, middleware/communication, framework/technical platform, task functionality, architecture, stimulus material, visualization provided?, timing, code for task provided?
Participants	recruitment strategy, incentives, age, gender, occupation, inclusion or exclusion criteria, approval of ethics committee
Experiment flow	experiment structure, instructions, training procedure, trial ordering, repetitions, blocks & breaks, pre-study screenings, questionnaires
Data processing	derivation of labels, data transformation, filtering, windowing, artifact cleaning, hyperparameter optimization, outlier handling, feature extraction, feature selection, learning model, evaluation procedure, processing code provided?
Brain signal integration	brain input effect, type of integration

in accordance to the typical section structure of the papers. This process resulted in the following *categories*: Technical aspects of recording, Task description, Participants, Experiment flow, Data processing, and Brain signal integration. Of course, such an experiment model contains aspects which may also be covered in a general model of experiments in HCI or cognitive psychology; however, we opted for creating a model specific to brain signals in HCI to cover the unique combination of experimental work in an HCI context together with aspects of cognition, neuroscience, signal processing, and machine learning. In the future, the model could be extended to cover other types of physiological interfaces.

In Table 3.4, we introduce these categories and list the attributes contained within them. Tables A.1-A.6, in the Appendix, provide further details, including definitions and examples. For each attribute, we give a definition of the item and present an example taken from the surveyed literature. We tried to identify examples which are biased towards a more detailed documentation, but individual examples may still lack certain information.

3.7 Reporting Practices in HCI Research with Brain Signals

In this section, we discuss our analysis to determine what aspects of an experiment the HCI community considered relevant to report, what structure we could identify through statistical analysis (compared to the a-priori structure of the experiment model), and what differences we encounter in relation to the variance of research as observed in Section 3.5. For this purpose, we annotated each of the 110 papers, as described in Literature Database Curation according to the presence or absence of each attribute in the presented experiment model (Section 3.6). Thus, we not only extracted statistics on *what* was investigated (Section 3.5), but also *how* it is reported. We then explore the prevalence of different attributes across papers and the relationships between different attributes and topics.

The goal of this review was not to judge specific papers for absence or presence of certain details, but to get a picture of how experiments are reported today, and what differences we observe in style of reporting for different types of papers. This provides an indication of what aspects of the experiment model are central to most HCI publications, what parts need refinement, and what parts are considered niche.

Attributes could be rated as present, absent, or not eligible (for example, "brain input effect" for non-real-time studies). Every paper was examined by one to four raters (median is 2), primarily authors of the paper and research assistants, all with background in human-computer interaction, computer science and/or neuroscience, and including graduate students, a senior postdoctoral researcher and professor. We first compiled a set of detailed definitions and examples for the different attributes and then all raters annotated the same paper and discussed the process to identify any questions or areas that needed additional clarity. Between raters, we arrive at an agreement rate of 81%. All raters could add additional information (e.g., additional attributes) or refinements to the definitions to the model during the annotation process. Ambiguities could be settled through a growing collection of examples and explanations. The annotated list of all papers as well as the data matrix of ratings is available as supplementary material.

3.7.1 Prevalence of Experiment Model Attributes

In Figure 3.1, we show the prevalence of the attributes of the experiment model across papers, i.e., the relative number of papers reporting the corresponding aspect. We see a large variety of resulting scores: A number of attributes related to data acquisition is present in nearly every paper in which they are eligible. These include mostly basic signal-related aspects, such as the type of device/sensor or sensor positioning. On the other end of the spectrum, we find that many attributes of data processing and machine learning are rarely reported, even if eligible.

One point to consider is that we rated attributes as absent in cases where no information was given, even if logically possible, because it was not considered by the authors. An example of such a case is the attribute "participant restraints." If the authors did not apply any restraints and thus do not report on this, this is indistinguishable from a situation in which a restraint was applied but not reported. A consequence of this observation is that it may be necessary to explicitly report attributes of the model as not applicable, to avoid such ambiguity.

To get a rough estimate of the impact of reporting these attributes, we calculated Pearson correlation between the fraction of reported attributes which were present in a paper with the number of citations per year since publication. This yields a modest, but significant correlation of $r = 0.16$ ($p = 0.04$), indicating that comprehensive experiment reporting may play some role in the paper's impact and success. Similarly, we correlated the number of reported attributes with the age of an paper, to study whether reporting changes over time globally. The correlation of $r = -0.12$ is not significant ($p = 0.12$). For a more detailed view, we repeated this analysis for individual attributes and found that some attributes, such as questionnaires and methods of synchronization, appear more frequently in recent publications (occurrence of these attributes correlates negatively with age), while others, such as participant restrictions or non-standard preprocessing techniques, occur less frequently in newer publications (occurrence of these attributes correlates positively with age). See Figure 3.2 for the attributes with the highest absolute values for r.

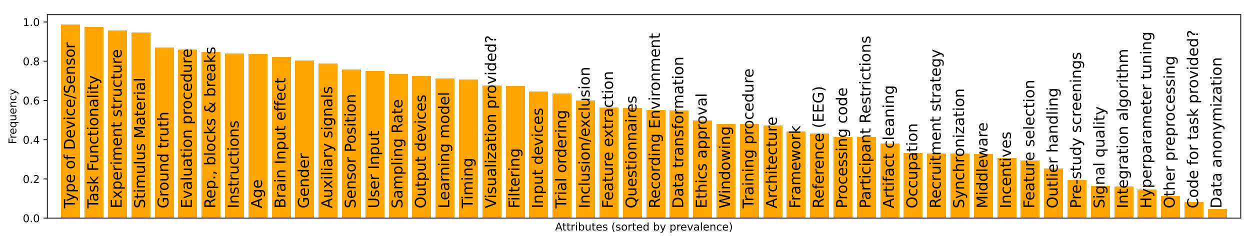


Figure 3.1: Prevalence of attributes across all analyzed papers, values indicate the relative number of papers reporting the corresponding aspect.

3.7.2 Structure of Experiment Reporting

In the following, we analyze the collected prevalence information on experiment model attributes to uncover a structure. While the proposed experiment model in Section 3.6 is structured based on expert perspective, we sought to explore whether a data-driven approach would reveal meaningful groupings that might support or refine that structure.

As a first step towards this goal, we investigated the (in)dependence of the different attributes of the experiment model to identify possible relationships or redundancies. We calculated the Pearson correlation between all pairs of attributes across all scores. On av-

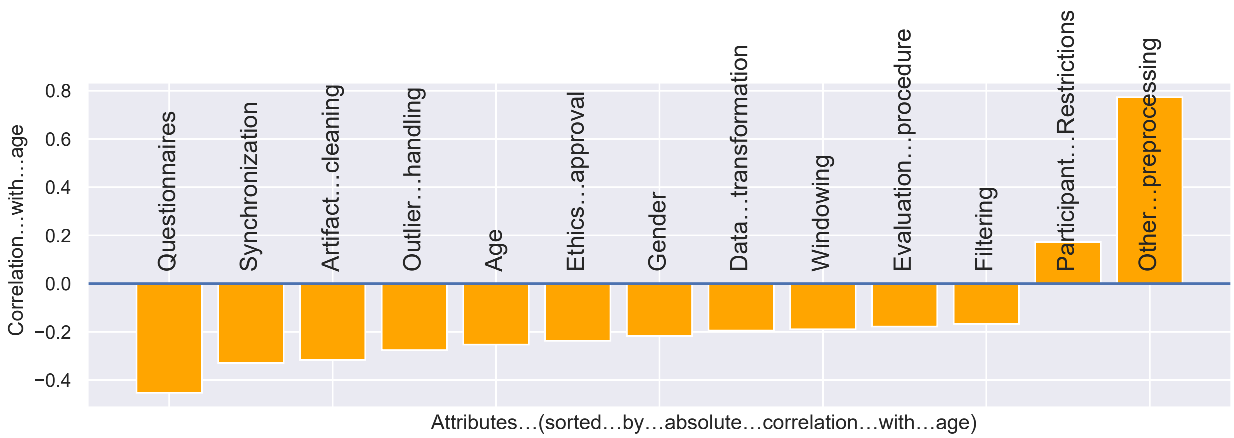


Figure 3.2: Correlation of model attributes prevalence and paper age for selected items, sorted by correlation coefficient.

erage, the maximum correlation between any attribute and others was 0.56, suggesting that most attributes are sufficiently distinct. However, some attribute pairs exhibited higher correlation (the highest value, 0.89, was between "gender" and "age" in the "participants" category), indicating some jointly occurring model aspects exist.

We explored this further by conducting an exploratory factor analysis with Varimax rotation on all attributes to investigate latent structural relationships underlying the analyzed papers, potentially revealing different reporting patterns for different types of publications. Factor analysis reduces the dimensionality of the data by grouping related attributes while preserving as much variance as possible. The factor analysis therefore allowed us to identify, through a data-driven process, a structure in the form of groups of experiment-reporting attributes that occur jointly in many papers. To preserve the variance of the original data, we avoided any scaling of variables or other preprocessing which could influence it. We summarize key points of our analysis and results below, leaving more detailed discussion in the full manuscript.

We found that ten factors were sufficient to explain the variance in our data. Examining the factor loadings—a measure of the correlation between a factor and attributes of our model—reveals that several of these factors mapped neatly to individual model categories, while others revealed unique groupings spanning multiple categories. For example, individual factors clearly mapped to the "participants" and "experiment structure" categories, while however, other factors mapped to clusters of attributes that were more broadly related to aspects of, e.g., experiment design or signal processing techniques. These clusters offer a complementary, data-driven structure that reinforces some elements of the original model while reflecting associations in reporting patterns. Based on the attribute loadings of each factor, we assign semantic categories that summarize the associated model aspects: F1: Fundamentals, F2: Experiment Design, F3: Data Processing, F4: Machine Learning, F5: Integration, F6: Software, F7: User Input, F8: Data Acquisition, F9: Quality Assurance, F10: Recording Conditions.

Finally, we used this data-driven categorization to investigate whether there were patterns in reporting dependent on the demographic information related to each paper (as in Section 3.5). We observe that papers which had different study designs tended to prioritize reporting different experiment details, with the type of brain signal integration (e.g., neurofeedback, explicit control, open-loop, etc.) emerging as a stronger differentiator than either contribution type or equipment used. This indicates that this alternative categorization is sufficient to meaningfully capture the breadth of HCI research with brain signals.

3.8 Expert Perspectives and Review

In Sections 3.6 and 3.7, we explored a model for the description of empirical work in using brain input with HCI methods and explored the current practices in existing literature. To understand what information researchers find to be relevant for reproducing and reusing work, we distributed an online questionnaire to experts and collected their feedback and comments. This is described below.

Questionnaire Structure

First the questionnaire asked for background information about the participant. Then, it asked questions to understand the extent to which the participant felt that their own papers were reproducible, followed by questions about their own sharing practices. The questionnaire then asked about the participant's experience reproducing or reusing other published work. Finally, the questionnaire contained all of the potential experiment model attributes along with a description and example as in the Appendix. Participants were asked to consider three contexts: *experiment reproduction*, (i.e., re-performing the same experiment), *data reuse* (e.g., running different analysis on published data), and *artifact reuse* (e.g. extending a presented software framework with additional functionality). For each of these three contexts, the participants rated the level of detail needed for each item as one of the following: *Not Needed at All, Not Very Detailed, Moderately Detailed* or *Highly Detailed*, which have been coded as 0, 1, 2, or 3 respectively. Throughout, there were opportunities for the participants to provide free comments.

Participants

To recruit experts to the study, we contacted all authors of papers considered in the previous sections. We received feedback from 12 external experts (from 6 different countries, ranging from graduate students to full professors from HCI and BCI) in the field who commented on the model and provided suggestions for minor adjustments. Two participants did not complete the entire survey, and we include their ratings and responses for the questions that they did answer.

Data and Code Sharing Practices

Four out of ten of the participants regularly share recorded data, data processing code, experiment processing code and study materials. Seven out of ten felt that most of their published studies are reproducible from the information given in the article. Space limitations were identified by four participants as the barrier to publishing fully reproducible experiments. In addition, one participant mentioned that the data is often regarded as medical data that has privacy requirements and cannot be shared and another participant mentioned that permission from the participants may not have been given for sharing. Another participant discussed the additional time requirements for preparing data for sharing. One participant mentioned that application or data processing code might not be shared for intellectual property protection.

Experience with Reproducibility and Reuse

We also asked the participants to reflect on any experiences that they have had reproducing study results from published work by other researchers. All six of the participants who indicated that they have attempted this, expressed that doing so was challenging and time consuming in some way. One mentioned getting lower accuracy than reported in the original paper. Another mentioned that lack of documentation about channel positions, markers, etc., made it difficult. One response directly contrasted the neuroscience/BCI/machine learning journals and the HCI publications, stating that the former is "much more reproducible" than the latter. Specifically, this participant mentioned that the former provides more details on algorithms, parameters, and precise timing of the protocol, while the latter is often, but not always, "vague/generic". They also pointed out the difference between practices of sharing code with "BCI/machine learning papers do[ing] it much more often (but not enough)."

Research Community Views on Reproducibility

We also asked participants to rate the level to which reproducibility is valued in their main research community. Six participants rated it either 4 or 5 out of 5 (high). Of these, one researcher identified their main fields as perceptual graphics and the other five identified BCI as their main research area. The four other participants rated this as 1 or 2 (not valued highly) and they identified their primary field as human-computer interaction, cognitive systems or computer security user behavior, with one mentioning it being "valued in theory, but no way of receiving credit for it".

Feedback on Experiment Model Attributes

The participants provided 41 comments to the different model attributes, leading to an incremental improvement of the category definitions. Their responses showed they rated the attributes of the model with an average of 2.04, with no model category scoring lower than 1.92. This indicates that the model lists relevant information. Bonferroni-corrected pairwise Wilcoxon signed rank tests showed that their importance ratings differed significantly depending on whether the experts rated the attributes for the purpose of *experiment reproduction*, (median importance of 2.26), *data re-use*, (2.12), or *artifact re-use*, (2.0). This shows that the context and viewpoint of the author and reader influence the priority for different attributes.

Summary

Many of the comments about experiences with reproduction and reuse of their own work as well as that of others aligns and confirms the motivation of this work. In addition, their feedback on the experiment model attributes constitutes a preliminary evaluation of the model.

3.9 Discussion & Outlook

From our observations in the literature survey and the experiment model, as well as the insights from the expert questionnaire, we derive a number of recommendations.

Exchanging Best Practices

Our analysis showed that different types of papers use different practices and could benefit from exchange. For example, we showed a divide between application-oriented and method-oriented HCI research in the types of sensor headsets used and the important trade-off between user comfort and signal quality. This divide becomes especially important if we consider that a substantial number of papers do not report details on aspects related to signal processing, sometimes taking the output of the commercial systems at face value. On the other hand, the method-oriented community could learn from others that comfort and user experience matter and working towards that is a merit which may even warrant a certain loss of classification performance (but not empirical validity). There is an opportunity for these papers, that have different end goals, to be written to have broader appeal and potential for reuse. Another cultural difference lies in the way that the modeled mental states and processes are referred to, i.e. from a neural perspective or a cognitive perspective. A common vocabulary and a better awareness of similarities and relations could help to better leverage common resources and insights.

It would also be valuable to turn specifically to the publications of the core BCI community to transfer some of the relevant standard practices of these experts in experiment design, signal processing, machine learning, and the underlying neural processes to the HCI community, to avoid common pitfalls, especially if they are BCI-specific. In particular, Brouwer, et al. describes six recommendations related to BCI studies, which include 1) define your state of interest and ground truth 2) connect your state of interest to neurophysiology, 3) eliminate or address confounding factors, 4) practice good classification methods, 5) provide insight into classification results and 6) justify the use of neurophysiology [35]. Jeunet, et al. details best practices related to signal acquisition, data processing, experiment design and the user component [157]. These BCI-focused papers make recommendations that can carry over into HCI research with brain signals.

HCI does not have a consistent record of meta-research and model-building as has helped the BCI field, (e.g. through the BNCI 2020 initiative), to establish common terms and public data sets. Besides improving internal communication, it would also help to make the field more accessible to HCI researchers who do not yet use brain signals in their research but could benefit from it.

Awareness of the Heterogeneity of HCI Research with Brain Signals

As authors, reviewers, and editors in the field, it is important to be considerate of the heterogeneity of HCI research with brain signals. To study and organize reporting practices in breadth, we first analyzed range of domains, of cognitive vs. neural states, of types of headsets, types of brain input integration, etc., to get an understanding of the diversity of concepts a model needed to cover. This "demographic heterogeneity" directly impacts experiment reporting. For example, the choice of headset determines whether details about sensor positions, filtering, data quality can be reported, and the factor analysis reveals differences in reporting for different types of integration.

Our analysis provides some evidence that the HCI research community using brain signals is different from the traditional BCI community. Furthermore, we showed that even within the community, there are different sub-communities with different reporting styles. This observation is not only relevant for future authors seeking guidance, but also for reviewers, who may use our findings to see what level of detail can be expected on the one hand, but also to recognize the range of acceptable contributions, on the other hand. Because of the diverse disciplinary backgrounds in the field, a critical challenge in publishing is the misalignment of expectations (e.g. papers being criticized for being "too neuroscientific" for an HCI conference, or papers "lacking rigor" when using commercial devices). Another source of heterogeneity is the wide spread in the acceptable number of study participants. Interdisciplinary research then runs into the risk of being "shot down" by multiple disciplinary reviewers, pulling the paper in multiple, opposite directions. This may discourage attempts at interdisciplinary work, which we consider crucial for the long-term development of the field. By uncovering the model that represents current HCI research with brain data, based on existing peer-reviewed publications, further discussion can emerge about expectations, reproducibility, and opportunities to strengthen future research in the field.

This recommendation should not be understood as a call for a lack of carefulness or willingness to improve. On the contrary, the diversity in research creates ample opportunity for learning for everybody, as elaborated in the previous recommendation. For example, researchers need to be aware and properly address the frequent problems that can lead to overfitting and non-reproducible results, such as improper handling of artifacts and premature parameter tuning. Utilizing strict validation criteria with pre-defined metrics, and appropriate chronological validation during offline analysis as well as the use of established processing pipelines can prevent that. However, not all learning is actionable immediately, e.g., because data has already been collected or necessary collaborations with experts from other areas need to be build up. We therefore think that in many cases, the inclusion of a transparent and honest discussion of shortcomings and opportunity for improvement may be an chance to offset imperfect, but thought-provoking research that others can build on (see also the third recommendation below).

Creating Opportunities to Build on Each Other's Work

Our analysis has shown that no publication or series of publications by the same group of researchers covers all aspects in the maximal depth. A reason for this is that complex research on brain input often involves multiple real-time components for signal processing, machine learning, and user interface design and no individual researcher is an expert in all of these areas. We recommend leveraging the fact that the research community in this area is large (396 unique authors contributed to the papers we analyzed) and diverse. This goal has several implications: First, it is necessary to provide enough information for others to be able to build on the work. Again, the experiment model can help to provide a sufficient level of detail. Second, we also recommend moving away as a community from expecting papers with full end-to-end systems as well as large-scale user studies all as novel contributions. A publication can have merit if it provides the research community with an exciting new experimental paradigm, even if the performed user studies has limitations. Incremental research which reproduces and advances an already published system can strengthen validity of results and ultimately leads to more mature artifacts. As an important support for this, researchers should be able to receive credit beyond citation of publications. Platforms like Open Science Framework (OSF) or Zenodo allow the assignment of DOIs to artifacts (software, data, or technical descriptions) which are not peer-reviewed papers and these can be referenced in a work that makes use of them. Finally, it should be noted that building on existing work can also mean to follow-up research on own studies, for example to confirm the results of an exploratory experiment through validation with independent metrics (e.g., validating a system for workload classification through a user study measuring efficiency).

Data and Code Sharing

One way of reporting experiments which seems to be still underused in the community is the sharing of data and code for analysis and experimental paradigms. It is important to further explore the challenges, such as the additional effort it takes to prepare the material with little perceived benefit or recognition. In addition, there may be concerns of increased scrutiny or loss of competitive edge in an academic race for high-impact publications. An "open" culture of publishing material could and should lead to a paradigm change, allowing the publication and joint improvement of research contributions: When an expert in empirical HCI develops an innovative experimental paradigm but an expert in biomedical engineering identifies areas to improve the published data preprocessing pipeline, both should be able to get credit for an updated joint work in contrast to work that is never published.

To move toward increased data sharing in the HCI community, however, there is a need for additional author and reviewer guidance and support. In some venues, particularly the main research conferences, anonymous review is expected, discouraging authors from sharing external links that may de-anonymize them. However, the paper submission systems often do not easily support the submission of data, code, metadata, etc. An author submitting a paper might not know the best way to anonymously share their data for review. Similarly, reviewers need guidance about what level to examine the data, code and other materials during the review process.

Publishing Expanded Experiment Descriptions

Another recommendation for HCI papers with brain signals is to consider publishing expanded experiment descriptions beyond the conference or journal paper. The analysis of the papers shows that no individual paper achieved more than 72% of coverage of all eligible model aspects. The missing details limit the ability for other researchers to easily build on the work. For example, a paper by Jones, et al. [161] is detailed in its reporting of technical aspects during EEG recording. Many newer papers omit many of these details. However, the experiment from 2003 is also relatively simple compared to modern experiment designs, such as [10], which in turn covers other aspects. The analysis shows that journal publications exhibit on average a better coverage of experiment model aspects compared to conference proceedings. This is likely due to the limitations in space, as brought up in Section 3.8 by study participants. We therefore recommend to make use of the opportunity to publish supplementary material as offered by many conferences and journals. In addition, publication in open data repositories gives the opportunity to share concrete information.

Both approaches allow to provide unlimited and more technical information. Alternatively, authors may consider accompanying blog posts or other forms of documentation beyond the publication manuscript itself. Future work needs to study obstacles for making data available [348], especially such reasons which are specific to brain data; one example is the question of long-term anonymity of brain data, that has implications on the ability to share such data liberally in the context of data protection and privacy laws.

Applying the Experiment Model

One outcome of the experiment model is that researchers can now compare past and future HCI publications against the model to identify strengths and weaknesses in reporting. The model was cross-checked through the analysis with 110 HCI publications to cover a wide range of information that is considered relevant to the HCI community for experiment reporting. It can therefore act as a reference for deciding what information to present in a publication or experiment description. This can help researchers to make their research more accessible and reproducible. We also recommend the consideration of aspects of the model which did not apply to your research, but for which there may be uncertainty if unreported. An example of such a situation is "participants restraints": If a researcher did not restrain participants in any way, it may not occur to them to report this in the paper; however, other researchers may wonder if no restraints were applied or whether they were not reported. Furthermore, our database can help researchers identify "role models" for certain types of experiments and certain aspects of experiment reporting. Finally, the experiment model can also be used by reviewers by providing a check list to guide the systematic and objective assessment of experiment reporting.

Further Refinement and Expansion of Experiment Model

We have presented an experiment model as a starting point for further discussion, critique, and extension of the model, as well as identification of the importance of its attributes for different use cases. One source for improvement might be an alternative structure or even a multi-dimensional structure, as the statistical analysis of the papers showed multiple potential approaches to group the model attributes into categories.

Despite being based on 110 papers, the model cannot cover the full breadth of all relevant papers and future use cases. For example, the database currently only contains one paper of multi-user BCI and thus does not reflect the specific aspects of such BCIs, for example how the different users communicate. Multiparty experiments are complex and may require the experiment model to cover additional aspects (e.g., in the roles of the different participants) as these designs become more common. As other novel study designs emerge, the experiment model would need to adapt to represent these new paradigms. Another possible area of refinement is the description of statistical analysis and the discrimination of exploratory and confirmatory aspects, which could be formalized beyond our analysis once an explicit discussion of these aspects becomes more common. Another limitation of our current analysis is that we did not consider papers in domain-specific publication outlets, such as neuroergonomics or learning research, as we wanted to focus on HCI venues. From the identified core domains (see Characterizing HCI Papers with Brain Signals), we can in the future extend the analysis to publications in these areas.

Finally, the model could also evolve more in the direction of a broader contextualization of research, documenting how researchers discuss their specific experiment in relation to ethical concerns and societal implications [87, 150, 149, 174] revolving around the use of brain data in HCI. Several specialized works present a systematic discussion of these aspects, and list the following characteristics of ethical issues related to BCIs: *personhood*, *stigma*, *autonomy*, *privacy*, *research ethics*, *safety*, *responsibility*, and *justice* [40, 60]. However, they also note that the discussion of ethical aspects is not widespread in empirical papers. This is also reflected in our database: Apart from *informed consent*, which is the most common one and already covered by the experiment model, we selected *autonomy*, *safety* and *privacy* as the most tangible ones from the list above and identified the papers which referred to them. When accounting for false positives (e.g., a discussion about "autonomous driving"), we found that 11.8% of all papers referred to safety (for example in reference to safety-critical BCI-applications), 6.3% referred to privacy (often in the context of the data collection), and 5.5% referred to autonomy (for example in the context of users with disabilities). This count is also generous, because sometimes the concepts are only mentioned briefly in the introduction or discussion.

3.10 Conclusion

In this chapter, our motivation was to characterize the current state of affairs to facilitate future discussion on reproducibility and reuse of HCI research with brain signals. We studied the diversity of HCI research using brain signals, with regards to domains and applications, modalities, measured mental states and processes, and more. From 110 publications since 1996, we showed the large variety and thus the broad applicability of brain activity measurement to improve or quantify HCI. The analysis also revealed that the studied field is heterogeneous and composed of sub-categories, which use different ways of reporting their experiments. We conclude that these differences may pose a challenge to understand, reproduce, and build on this research.

One result of this work is the creation of an initial experiment model, which acts as a unified superset of recurring aspects of empirical work on using brain signals for HCI purposes. This model and the example attributes can act as a guideline to structure a report on empirical work in this area. It can help to reduce the mental workload and uncertainty for both authors and reviewers by providing structure for publications. Explicating, structuring, and naming different aspects of the model can facilitate a discussion in the community on what and how to report. The empirical analysis has confirmed the general structure of the proposed model but also revealed potential alternatives for future explorations. We saw that while some parts of the model define a minimal standard most papers report on, there are other aspects that are only addressed by parts of the community. A discussion and re-iteration in the scientific community would be an important follow-up from this starting point to create a more mature and comprehensive version of the model. This requires a broadening of perspectives beyond the scope of this analysis, e.g., through a workshop series at one of the premier HCI conferences. We made the model accessible through a GitHub repository at `https://brain-signals-hci.github.io/experiment-model/`, through which interested researchers can send pull requests to submit suggestions for improvements.

While there is much room for subsequent research, the presented work is an initial step toward reproducibility and reuse of HCI research with brain signals.

Chapter 4

BrainEx: An Easy-To-Use Tool for Exploring Brain Signal Datasets

Abstract

Technology advances and lower equipment costs are enabling non-invasive, convenient recording of brain data outside of clinical settings in more real-world environments, and by non-experts. Despite the growing interest in and availability of brain signal datasets, most analytical tools are made for experts in the specific device technology and have rigid constraints on the type of analysis available. We developed *BrainEx* to support interactive exploration and discovery within brain signal datasets. *BrainEx* takes advantage of algorithms that enable fast exploration of complex, large collections of time series data, while being easy to use and learn. This system enables researchers to perform similarity search, explore feature data and natural clustering, and select sequences of interest for future searches and exploration, while also maintaining the usability of a visual tool. In this chapter, we describe the distributed architecture and visual design for *BrainEx*, and illustrate its improved performance relative to existing systems. Additionally, we report on a preliminary user study in which domain experts used the visual exploration interface and affirmed that it meets the requirements. Finally, it presents a case study using *BrainEx* to explore real-world, domain-relevant data.

4.1 Introduction

Recent innovations and declining costs for non-invasive brain monitoring technologies are paving the way for future innovations in brain-computer interfaces, clinical applications, and intelligent systems that adapt to changes in an individual's dynamic cognitive state [94, 310]. While existing tools help with brain data acquisition and signal processing, most are geared toward biomedical engineers, scientists, or clinicians who analyze the sensor data

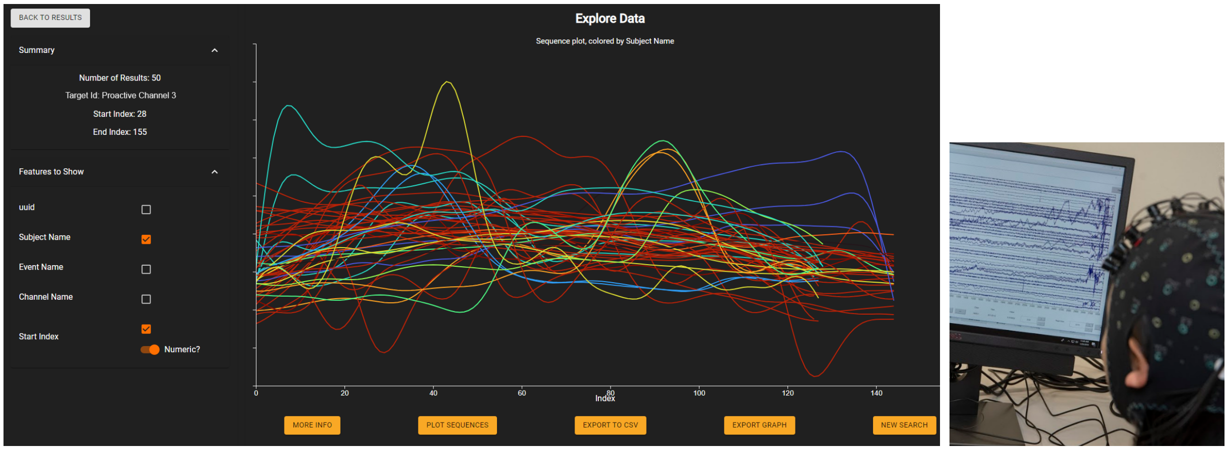


Figure 4.1: BrainEx is a web-based visual analytic tool, designed for exploring sequence similarity and clusters within brain signals. On the left, we see 50 similar sequences with color used to encode metadata about the search results, and on the right is a person wearing a functional near-infrared spectroscopy brain sensing cap.

in highly controlled and specific contexts and are experts in the underlying device technology. Their well-established practices come from traditional neuroimaging and neuroscience studies where data is collected following strict experimental protocols and analyzed by the same team that collected the data. While these practices have enabled groundbreaking insights into the brain, they are less effective in more real-world contexts or when datasets are reused by others.

In these contexts, exploration can be valuable for gaining familiarity real-world data and identifying brain signal patterns that might indicate a cognitive state of interest. For example, researchers or developers may be interested in finding signature signal patterns that indicate that a driver is distracted, or a student is focused. They might also want to scan datasets for recurring patterns that suggest underlying events or mental processes common across conditions or experiments. Researchers may also be interested in finding events or tasks that lead to similar brain signals, even if they had not been associated together prior to the data collection (e.g. doing a particular math problem or detouring while navigating). These exploratory steps could inform future confirmatory studies. However, current tools are not designed to support this kind of data-driven exploration and do not leverage advances in time series data mining.

In this chapter, we introduce *BrainEx*, a web-based, brain data analytics platform for visual exploration and discovery within time series datasets. Its core design philosophy is *exploration at every stage*. *BrainEx* builds on data mining approaches for interactively finding similar sequences in large datasets, and integrates them into a workflow specifically designed for brain signals. We focus on signals from functional near infrared spectroscopy (fNIRS), a non-invasive neuroimaging tool [339, 48] that has been used to measure cognitive states in real time [3]. *BrainEx* preprocesses datasets, uncovers structural patterns, and enables users to interactively query, filter, and visualize similar signals, supported by metadata and contextual information. This way, *BrainEx* is designed not simply for statistical analysis, but broadly for empowering users to better understand brain data, and to explore it in search of meaningful relationships for further study.

4.2 Contributions

This chapter presents an abridged version of the full BrainEx manuscript [139]. Readers interested in further implementation details or performance benchmarking should refer to the original publication. Here, we detail the following contributions, focusing on the system's functionality and accessibility to non-expert users:

- *BrainEx* provides an interactive visual interface for brain signal exploration, supporting cluster browsing and query-by-example similarity search across datasets. The use of multiple similarity metrics allows researchers to uncover patterns that might otherwise remain hidden.

- A user study conducted with experts in data visualization, neuroscience, and human-computer interaction demonstrates *BrainEx*'s effectiveness at achieving five functional goals: similarity search, feature distribution exploration, cluster exploration, integration between different components of *BrainEx*, and accessibility to researchers from different backgrounds. The positive feedback shows promise for advancing the field of brain-computer interaction.

- A case study with real-world fNIRS data illustrates how *BrainEx* supports discovery of relationships between task events, brain regions, and cognitive states, highlighting its potential as a tool for exploratory research.

4.3 Background

The *BrainEx* system brings together research on brain sensing with research on time series data mining and visualization to address challenges in brain signal analytics, with a focus on functional near-infrared spectroscopy (fNIRS) used in HCI settings. This section provides background on fNIRS, analytics tools for fNIRS, as well as time series similarity search and clustering.

4.3.1 Functional Near-Infrared Spectroscopy (fNIRS)

Functional near-infrared spectroscopy (fNIRS) is a noninvasive form of neuroimaging that provides time series data about cortical hemodynamics, which is correlated with brain activity [265], using near-infrared light. fNIRS relies on the fact that infrared light can penetrate human skin and is absorbed in different amounts dependent on the oxygenation of the blood. An fNIRS cap (Figure 4.1) contains multiple fNIRS light sources and light detectors on it, with each source-detector pair forming a *channel* of brain data. These measurements allow researchers to compare activity in different areas in the brain at the same time [94]. fNIRS is a useful tool for researchers due to its accurate, non-invasive, and portable properties. fNIRS research is a growing field within neuroscience [94]; in 2020, approximately 500 papers were published on the subject [131]. In addition, there has been a growing number of HCI publications that use fNIRS brain signals [368, 214, 209, 21, 269, 215, 158, 355, 231, 11, 67]. [214].

The data generated from an fNIRS study typically consists of multivariate time series, with one time series from each channel in the fNIRS cap (e.g. [355]). These time series represent the oxygenated and deoxygenated hemoglobin in the location where the channel is placed on the head. In addition to the fNIRS data itself, a dataset usually includes additional metadata that describes the sensor locations, the study participant identifier, the activity or events that occur during the study, among other things. These characteristics are similar to other brain sensing modalities, such as electroencephalography (EEG), as well as physiological sensing channels. Only the shapes of the signals and the sampling rate would differ, depending on the sensing modality.

From these multivariate time series and the associated metadata, researchers typically search for patterns in the brain data that indicate a particular cognitive or emotional state. This can be done by using statistical methods to look for significant differences between two sets of labeled sequences (e.g. distracted vs. focused). These labels would come from the experiment design where particular states are elicited in a controlled way and then marked or labeled in the data. Machine learning approaches are also common where labeled data is used to build a classifier for future unknown data. Both statistical and machine learning approaches can be difficult when brain signal sequences are of different lengths or different scales, but workarounds exist.

4.3.2 Analytic Tools for fNIRS

With the growing field, specialized tools have been developed to aid researchers in the analysis of fNIRS brain data [112, 208, 146, 319, 137, 373, 292] and each device typically comes with some basic analysis software (e.g., NIRx NIRSLab). These tools generally support the calculation of oxygenated and deoxygenated hemoglobin, as well as various filtering, signal processing, and visualization techniques [137, 373, 146]. However, many of these tools are specialized for specific experiment designs or application areas, typically supporting task-based (HOMER [146] and POTATo [319]), resting-state (NIRS-KIT [137]), or real-time analysis (Turbo Satori [208]). While widely used in for analyzing fNIRS data, these tools were primarily designed for domain experts and do not prioritize visual exploration, a key feature for other medical fields [42, 43, 44]. Few options exist for interactively exploring large or unfamiliar datasets, especially in the context of functional near-infrared spectroscopy.

Facilitating brain data exploration is particularly critical today given the growing emphasis on sharing diverse datasets of increasing size and complexity. This means that researchers may look at brain signal data that other researchers collected, and that they are not necessarily familiar with. As more fNIRS data becomes publicly available, new insights may be unlocked by applying time series data mining techniques and drawing from related work in other domains. However, effectively analyzing heterogeneous datasets remains a challenge.

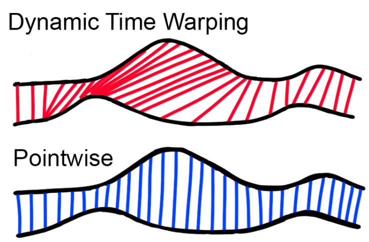


Figure 4.2: Comparison of how DTW and point wise similarity matching occurs. DTW (on top) allows for a one-to-many mapping, as seen by points from the top sequence all mapping to a single area on the bottom sequence that occurs earlier. Pointwise (on bottom) only allows for a one-to-one mapping of points that occur at the same time in both sequences.

4.3.3 Exploring Similar Sequences in Time Series

Finding similar sequences of fNIRS data is an essential operation for identifying brain signals that might be indicative of key cognitive or emotional states. Similarity search also plays a prominent role in many other areas, such as finance or meteorology. Like brain signal data, datasets from these domains can be large, often spanning tens of thousands of time points at minimum [242]; thus existing work on efficient similarity search for large datasets informs our approach.

Pointwise distance metrics, such as Euclidean or Manhattan, are easy to compute, but are limited because they require sequences to be equal in length and aligned temporally, restricting their applicability to real-world datasets. *Elastic distance metrics*, such as dynamic time warping (DTW) [27, 282, 300], enable more flexible comparisons by compressing or expanding the time axis, allowing a single point in one sequence to map to multiple points of another (Figure 4.2). Thus comparisons via elastic distance metrics focus on the *shape* of sequences rather than their *values* [72]. DTW's flexibility has made it a popular choice for similarity search in several domains, including comparing RNA sequences in bioinformatics [2], ECG pattern matching in medicine [41], and matching temporally misaligned sequences in brain data [80, 51, 216]. However, its quadratic time and space complexity [164, 239] and lack of a proven triangle inequality make it difficult to scale to large datasets. Nonetheless, several modifications have been proposed to improve its efficiency [169, 92, 374, 340, 281].

Although DTW is widely used, typical implementations only support a small collection of distance metrics—usually Euclidean and Manhattan. Prior work has shown that different distance metrics are better suited to different domains [46, 241]. A poor choice of distance metric may distort similarities between sequences, leading to misclassification [241]. As a result, effective exploratory tools must support multiple distance metrics to ensure robust pattern detection across contexts [163, 306, 316, 320].

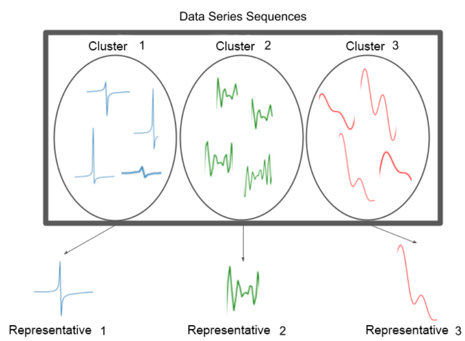


Figure 4.3: Representations and groupings derived by GDTW. The colors of the sequences correspond to clusters of similar sequences and their respective representatives.

Generalized Dynamic Time Warping (GDTW) [243] builds on this need by aligning sequences using diverse pointwise distances, allowing for more flexible and accurate matching (Figure **??**). GDTW employs clustering via simple distances (e.g., Euclidean, Manhattan) and refines results using warped counterparts like DTW or warped Chebyshev [243, 241]. Such flexible and generalized approach is a promising basis for interactive brain data exploration. However, there are no existing tools that support this for general use by non-experts.

4.3.4 Efficient Sequence Similarity Search Using Multiple Warped Distances

BrainEx rests on the foundation created by the frameworks introduced in ONEX [242, 244] and GENEX [243], enabling researchers to perform very fast and highly accurate similarity searches in large datasets. We briefly discuss these approaches below.

ONEX [242] introduced an offline preprocessing approach that clusters sequences using Euclidean distance, then performs similarity search using DTW. This leverages a customized triangle inequality between the distances to allow efficient querying while maintaining accuracy. By allowing for fast, accurate similarity search online, researchers may explore the similarity of time series more easily and interactively.

GENEX [241] generalized this idea by allowing researchers to warp their own distances and then incorporate them into the clustering pipeline, offering greater flexibility. However, it is limited by high memory requirements and long response times, which make it challenging to scale to large datasets. A distributed system would be better suited to provide interactive response times and reasonable memory requirements.

4.4 BrainEx Engine Architecture

When conducting studies with brain signals such as fNIRS, researchers generate large, complex and often noisy datasets. These datasets consist of multivariate time series of brain signals coming from multiple scalp recording locations. They also contain metadata documenting the participant, the sensor channel, and any events occurring during the experiment session. A common goal in this research would be to better understand the impact of one or more of these attributes on the corresponding time series data. For example, researchers may want to answer questions such as: *What parts of the dataset look the most similar to a particular instance of user distraction? Are there particular patterns that are frequent in the dataset, in general? Are there particular patterns that are frequent for a particular study participant? Or sensor location? Or event? Or a combination of these factors?* To answer these questions, researchers cannot assume that all sequences in the dataset are the same length, since real-world tasks can vary. However, the search results should still find the most similar sequence. Many of these questions could be answered by building on the foundation of GENEX.

In this vein, we created *BrainEx*, extending GENEX with a distributed preprocessing architecture that improves both speed and memory efficiency (Figure 4.4). *BrainEx* also leverages rich metadata to support flexible operations such as cluster exploration, filtering, and sorting through a user-friendly interface accessible to non-coders. Its modular design allows researchers to integrate new distance functions with minimal effort, making it more adaptable for diverse BCI research contexts.

BrainEx is implemented as a full-stack web application with three main components: a Python-based engine using Apache Spark for computation, a Django API, and a React front-end 4.5). The server handles dataset preprocessing and similarity search, while the browser-based interface enables users to explore clusters and view results. For implementation details, see the full BrainEx paper.

In the following sections, we briefly summarize the algorithms and distributed architecture underlying *BrainEx* which enable interactive visual exploration and discovery (Sections 4.4.1–4.4.2; further details in [139]). Section 4.5 continues with a description of the front-end interface.

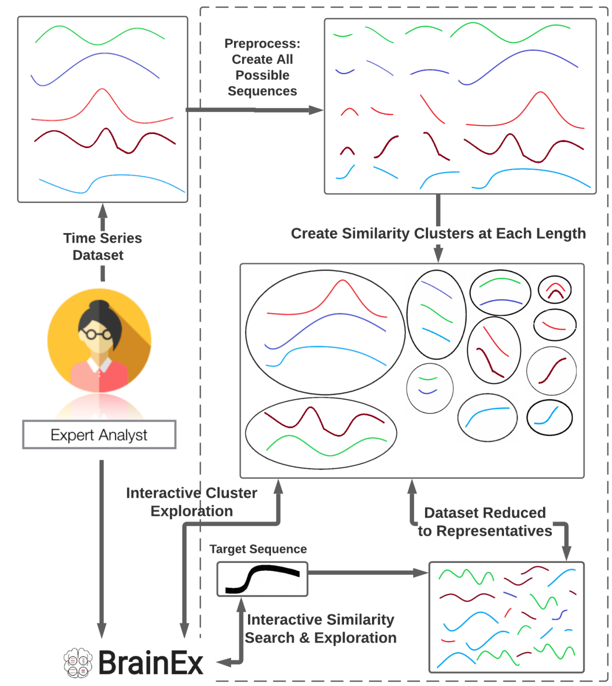


Figure 4.4: *BrainEx* Pipeline Overview. When preprocessing a dataset, the time series are divided into all possible sequences of all possible lengths and then clustered into similar groups of equal length. After preprocessing, researchers can interactively explore the clusters and perform a fast similarity search by finding the cluster representatives most similar to the target sequence, and only searching the clusters they represent.

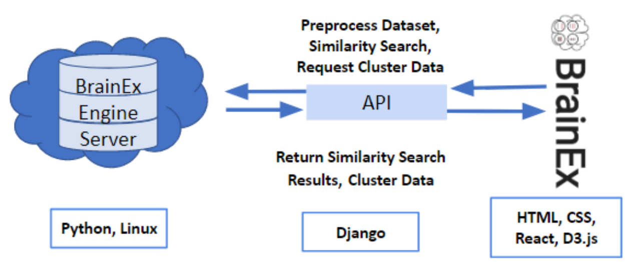


Figure 4.5: The BrainEx System Architecture. The first component on the left is the *BrainEx* Engine Server, developed in Python and usable with Linux OS. The middle component is the API that preprocesses datasets, performs similarity searches, and clusters data which is implemented with Django. The last component is the *BrainEx* website interface which the user accesses and is developed in HTML, CSS, React, and D3.js.

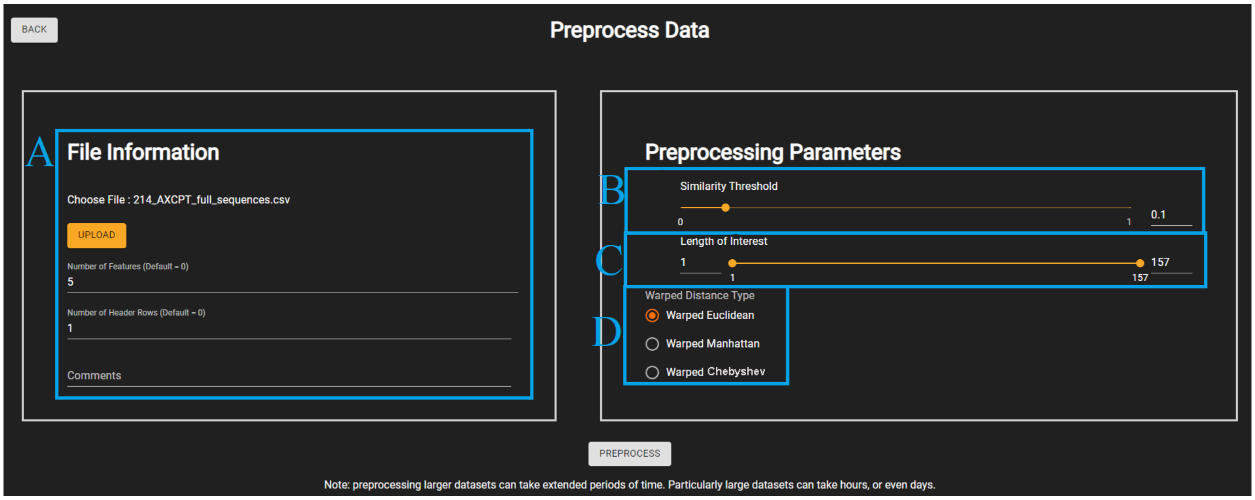


Figure 4.6: The Preprocess Data page is where a user will upload a new dataset and specify the parameters for preprocessing. **A)** Users must manually provide the number of header rows and feature columns; the default for each is 0. **B)** The user-defined similarity threshold is the minimum similarity requirement between sequences in the same cluster. **C)** The length of interest is the range for subsequences to be spliced; the default is 1-n where n is the maximum sequence length in the dataset. **D)** The available warped distances for similarity matching. Currently, Warped Euclidean, Warped Manhattan, and Warped Chebyshev are available, however the code is built to easily accept more warped distances.

4.4.1 BrainEx Engine: Distributed Preprocessing Algorithm to Compute Clusters

To enable fast similarity search, *BrainEx* organizes time series data into clusters based on sequence length and a user-defined similarity threshold (Figure 4.6). Each cluster contains sequences with similar structure and is represented by a single exemplar sequence. This clustering enables efficient querying by comparing against a smaller subset of representative sequences.

Improving upon the resource requirements of GENEX, the clustering process uses a distributed preprocessing algorithm that segments time series data across multiple computing cores using a balanced slicing method we call *Generalized Distributive Step Slicing* (Generalized DSS), shown in Figure 4.7. After segments are distributed equitably across compute workers, each worker independently forms clusters using efficient pointwise distance measures (e.g., Euclidean or Manhattan), checking whether a new sequence fits within the similarity bounds of an existing cluster or should form a new one. Cluster representatives are fixed so each cluster will have the first sequence that was sent to that cluster as representative, to simplify processing and preserve natural variation in the data. Preprocessing is complete once all workers have clustered their respective sequences.

At this point, the user can investigate each individual cluster to understand its characteristics, such as what features are present in the cluster, how many sequences it contains, and the length of the sequences. This feature will be referred to as *cluster exploration.*

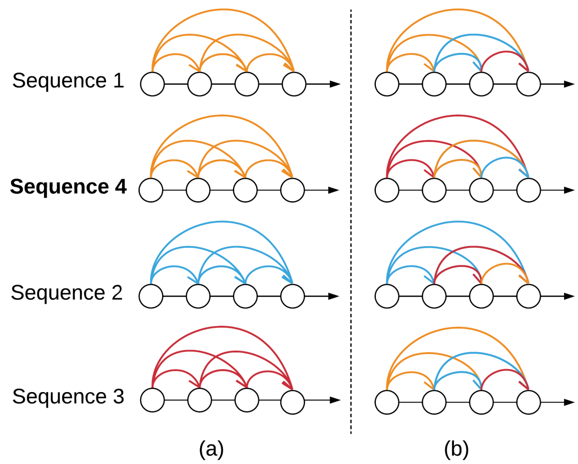


Figure 4.7: Parts (a) and (b) demonstrate the difference between the naive distribution scheme and our implementation of distributive step slicing (DSS). The discs on the time series are individual data points. The curves above the data points represents the sliced subsequences. Different colored curves represent work done by different executors. Figure (a) shows the naive distribution scheme, when the number of time series is not divisible by the number of executors. This results in load imbalances. In this case, the load of the 'orange' executor is twice the amount of its fellows. Figure (b) shows a Generalized DSS for load (number of sub-sequences) balances over multiple time series; note that the executors' start index for each time series is set in a round-robin style to ensure further balancing of the loads.

4.4.2 *BrainEx* Engine: Distributed Similarity Search Algorithm

BrainEx allows users to query datasets by example, returning sequences similar to a selected target. As with GENEX [241], instead of searching all sequences, *BrainEx* compares the query to cluster representatives to the user-specified similarity threshold. If a representative is sufficiently similar, the system searches within that cluster to find matching sequences. This reduces computational load while preserving accuracy. As with cluster formation above, this operation is performed in a distributed fashion to increase computational efficiency. It enables real-time interaction with large datasets, helping users surface relevant brain signal patterns across different time scales or contexts.

4.4.3 *BrainEx* Engine: Time Series Indexing and Memory Optimization

To reduce memory demands, *BrainEx* avoids storing all possible subsequences directly. Instead, it uses lightweight indexing with unique identifiers and start/end markers to represent each subsequence. During computation, these index objects are used to retrieve subsequences from the full dataset as needed, and are quickly discarded after use. This approach ensures that *BrainEx* can handle large datasets efficiently without overwhelming system resources.

4.4.4 *BrainEx* Engine: Operations

BrainEx performs three main operations: *best match selection*, *ranked similarity search*, and *cluster exploration*. To assist in these operations, the user is presented with a number of parameters to filter the results. Below are descriptions of the parameters and operations they assist; more robust descriptions of similarity search and cluster exploration are found in Sections 4.5.4 and 4.5.5 respectfully, where we discuss the user interface.

Similarity Search

BrainEx's similarity search includes four parameters to adjust either the query or the sequences that match the query. Users specify their *target sequence* from the preprocessed dataset and then specify the *start and end index* from that sequence to search for matches. The returned matches will be approximately the same length as the target sequence with a ± 1 margin of error. The user can choose any *number of matches* for their target sequence with an upper limit of the total number of sequences of similar length to the target. The last two parameters limit the sequences that can be selected as a match. First, the user can limit the *overlap* between results to prevent results from the same parent sequence that are offset by only a few data points. Secondly, the user can exclude sequences that include points from the target sequence.

Cluster Exploration

All clusters can be filtered by how many sequences are grouped in the cluster (*cluster size*) and the *length* of the sequences. Each cluster contains sequences of a single length. Therefore, if one sequence in the cluster is of length 40, all of its sequences have length 40.

In addition, the user can also filter clusters by values of specific *user-customized features*. In brain signal datasets (e.g., fNIRS), common features are *participant name*, *event name*, and *channel name*. By filtering by these features, the user can search for clusters that are primarily composed of a certain event or a specific subject, or look for clusters that have an equal number of sequences from different channels.

4.4.5 *BrainEx* Engine: Supported Datasets

BrainEx supports the input of TSV and CSV files. Columns contain features or datapoints for a sequence while each row is an individual sequence. Each sequence must contain the same number of features, but need not be the same length. We show below a snapshot of an example dataset, containing one header row, two features, and sequences of length 3.

Subject	Condition			
1	*Individual*	1.1	1.2	1.3
1	*Cooperative*	2.3	2.4	2.5
2	*Individual*	1.9	2.0	2.1
2	*Cooperative*	2.8	2.9	3.0

4.5 BrainEx Visual Exploration Design

The *BrainEx* engine described above provides sophisticated time series analytical tools to enable interactive exploration. All of the functionality described above can be accessed using a command line API. However, this requires domain expertise and deep technical knowledge to execute. In addition, without visual representation of the data, the functionality is not particularly useful for an end user to familiarize themselves with and truly explore the data.

Thus, we aimed to build an effective visual interface on top of the *BrainEx* engine to expose the underlying cluster exploration and similarity search algorithms in a way that is easy to use by users of all skillsets. The goal is not necessarily to introduce novel ways of visualizing time series data, but rather to make the underlying algorithms more accessible and valuable to researchers, so they can more easily explore and discover interesting relationships in time series datasets. Users of *BrainEx* may not be familiar with command line interfaces and would not be able to use *BrainEx* to its full potential without a visual interface.

4.5.1 Usage Goals

We developed the following simple usage scenario to motivate the design of the *BrainEx* interface. A researcher has performed an fNIRS study, collecting data from several participants as well as multiple sensor locations on each participant, creating multivariate time series signals. In this hypothetical study, participants were asked to complete several short tasks, some that were calibrated as *easy*, others that were calibrated as *hard*, and some task with *unknown* difficulty. The researcher is interested in better understanding this data during the task of *unknown* difficulty, and would like to use the data collected during the other two calibrated tasks to see if there are connections.

In this scenario, a researcher would need to get a sense of the distribution of the data by understanding which sequences are similar to each other in the dataset. In this case, they would not have any particular sequences of interest in mind. Instead, they would need to explore the entirety of the data. The researcher may want to understand how the sensor location and task are related to the underlying brain activation. They may also want to get a sense of the "shape" of a time series which represents some grouping. In the process of exploring, a researcher might come across a sequence which is of particular interest. It is important to be able to search for any number of sequences similar to the one discovered. Once these are retrieved, they will need to explore the distribution of the related metadata. When a researcher has identified any subset of the data, whether by exploration or searching, would need to be able to explore the distributions of one or more features of this data. This is just a simple example to help identify the functional requirements for the visual interface design, but much more complex analysis is possible; some of this is illustrated in the case study described in Section 4.8.

4.5.2 Functional Requirements

By exploring the usage goals, we determined five functional requirements to incorporate from the *BrainEx* Engine into the user interface. The first requirement is the ability to compare one sequence to the rest of the dataset by finding which sequences it is most similar to.

Requirement 1: Similarity Search

a) Support retrieving and ranking any number of sequences similar to another sequence of a researcher's choice.

b) Support exploration of search results and attributes of the sequences in the result set.

The second requirement is the ability to explore the feature distribution in a set of sequences that are naturally grouped together.

Requirement 2: Feature Distribution Exploration

a) Support exploring the distribution of a single feature in a cluster or result set.

b) Support exploring the joint distribution of two features in a cluster or result set.

c) Support comparing the relationships between three or more features in a cluster or result set.

d) Support identifying a sequence shape that well-represents a cluster or result set.

Similarly, *BrainEx* should support the ability to explore the distribution of all such natural groupings in the dataset.

Requirement 3: Cluster Exploration

a) Support exploring the range of cluster sizes and sequence lengths within clusters.

b) Support finding clusters of similar sequences with skewed feature distributions (i.e. clusters that mostly contain a particular User, Channel, or Event).

From the insights gained by exploration into clusters of sequences and feature distributions, *BrainEx* should support using these insights to start new explorations and searches on sequences found to be of interest.

Requirement 4: Integration

a) Support searching and finding sequences of interest to explore further based on the results of cluster exploration.

b) Support exporting sequences of interest and other findings to explore further in other tools.

In addition to the functional requirements above, the final requirement was that use of the tool should not be limited to a small group of highly trained researchers with access to high performance computing.

Requirement 5: Accessibility to All Researchers

a) Support fast computations, regardless of researcher's computer.

b) Support researchers at all levels, from novice to expert.

c) Support diverse experiments to be explored, and remain agnostic to the particular user-customized metadata (e.g. participant, events, channels, etc.) that are associated with the dataset.

4.5.3 Interface Components

Based on these requirements, we built a visual interface on top of the *BrainEx* engine. It provides an integrated pipeline for researchers to move between the broad exploration of a dataset and queries for specific sets of most similar sequences. Our tool is fully agnostic, and can accommodate any number of customized, researcher-specified labels on the data. A user can select any of the preprocessed datasets and then select *Similarity Search* (Figure

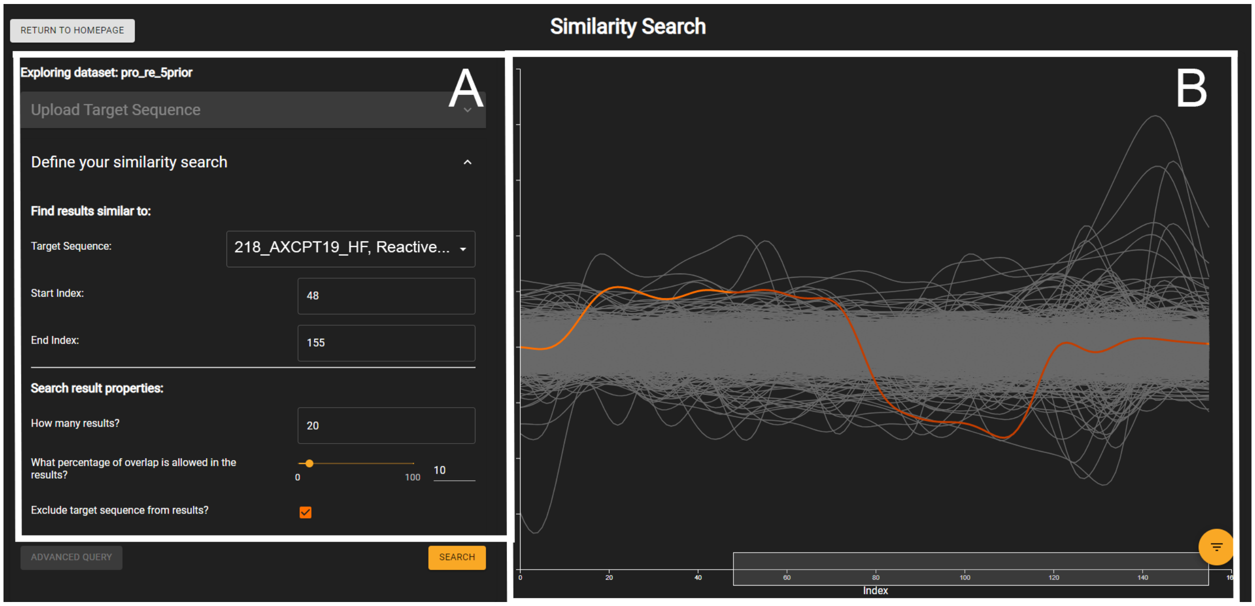


Figure 4.8: Similarity Search. *BrainEx* Visual interface for conducting similarity search consisting of search options in the left panel (marked A) and a visualization of the dataset's sequences on the right (marked B).

4.8) to find the best k matches to a specific subsequence of the dataset *or* the user can select *Cluster Exploration* (Figure 4.9) to explore the distribution of features among clusters in the dataset.

4.5.4 Similarity Search

Our visual interface leverages the power of time series analysis [242, 241, 244] and expands it with interactive visualizations, as well as analytic workflows developed for fNIRS data analysis, to provide the time series exploration experience outlined in Figure 4.4.

BrainEx enables researchers to visually investigate the sequences in the dataset. Before initiating a similarity search, researchers can explore the sequences in the dataset via a line chart. To reduce the number of sequences visible and to enable more targeted exploration, *BrainEx* provides filtering, zooming, and panning support in the sequence view seen in Figure 4.8.B. For a more specific search, *BrainEx* enables researchers to select start and end indices, specify a maximum number of results to return, and exclude the target sequence from the search results. As there may be overlap among similar sequences, *BrainEx* also enables the selection of a percentage of allowed overlap. These filters can be set with the options shown on the left side in Figure 4.8.A.

Once the search is completed, results are presented via a table containing details about the resulting sequences and a line chart where each sequence is visualized. The table enables the researcher to see each sequence, the features associated with it, and its distance from

Figure 4.9: Cluster Explorer. *BrainEx* interface for exploring the clusters consisting of a table of filtering options in the left panel (marked A), a table of the dataset's clusters and their color-coded feature distributions (marked B), and a visualization showing the overall distribution of clusters in the dataset (marked C). When a cluster is selected, the visualization changes to show that cluster's representative.

the target sequence. For easy comparison, the target sequence is always located at the top of the table and is highlighted in the line chart. Hovering over a sequence in either the line chart or the table highlights the sequence in both locations and scrolls to the sequence in the table if it is not already visible. To enable more refined control over the visualization of the results, the sequences can be sorted and filtered by each feature, the number of visible results can be limited, and a maximum distance can be specified. *BrainEx* also allows researchers to export the resultant sequences as a CSV file and to save the line chart as an image for further exploration and/or interpretation. Moreover, the distribution of features in similar sequences can be investigated through the feature distribution explorer.

4.5.5 Cluster Exploration

BrainEx enables researchers to explore the clusters of sequences through a table (Figure 4.9.B) containing information about every cluster in the dataset. This information includes the *number of sequences* in each cluster, the *length of the sequences* in each cluster, and the single *distribution of the features* of sequences in each cluster. Table cells are colored to show the salience of particular feature values, allowing users to scan for clusters with interesting distributions to investigate further.

The number of clusters can grow quickly as the number of sequences and the range of sequence lengths increases. To provide researchers with the ability to target their exploration in this large space, *BrainEx* provides sorting on each column, filtering, and a visualization. Researchers may use sorting to quickly find clusters that may be more interesting, such as clusters with more sequences or those with particularly skewed feature distributions. This sorting can also be tuned by using the filtering on the length and number of sequences in the clusters that are shown in the table (Figure 4.9.A). As researchers may not know how best to apply these filters, we also provide a scatter plot (Figure 4.9.C) showing the distribution of clusters against the number and length of sequences within each cluster. This visualization can provide context for the researcher's expectations when applying the filters, while also describing more general properties of the dataset's sequences and their tendencies to be clustered together.

In addition to exploring the distribution of clusters and the features within those clusters, researchers may also explore the shapes of sequences in each cluster. These shapes are provided in the form of the cluster representative sequence to which the other sequences in the cluster are similar. Once a researcher selects a cluster, they see a line plot of this representative sequence to show this shape with a scale to allow comparison between the representatives of different clusters.

4.5.6 Feature Distribution Exploration

Once *BrainEx* retrieves a group of sequences, either from a similarity search or from cluster exploration, it supports visualization of relevant information. This is complicated by the fact that *BrainEx* is fully agnostic of the user-defined, customized metadata provided. When researchers preprocess the dataset, they can specify any number of attributes (e.g. channel, user id, condition), which can have different values and data types. User-provided datasets may also vary in size.

As a result, the visualization techniques must be able to show the distribution for an arbitrary number of sequences. Further, they must be able to present the distribution with respect to any number of feature labels. Thus, *BrainEx* combines several data visualization techniques to present data depending upon the sort of information a researcher is looking for.

Researchers can choose the set of features that they are interested in by using feature specification checkboxes. As they check boxes, the visualization display pane updates in real time, allowing for seamless exploration. The display may change among four states: single feature bar charts, two feature heatmaps, many feature parallel coordinates views, and a time series sequence view. Detailed information about individual visualizations is presented when a user clicks the "More Info" button.

The first visualization state is a bar chart (Figure 4.10.A), which is used whenever a researcher needs to display the distribution with respect to one feature. Bar charts are well-studied for the application of comparing two or more values in a single dimension [57].

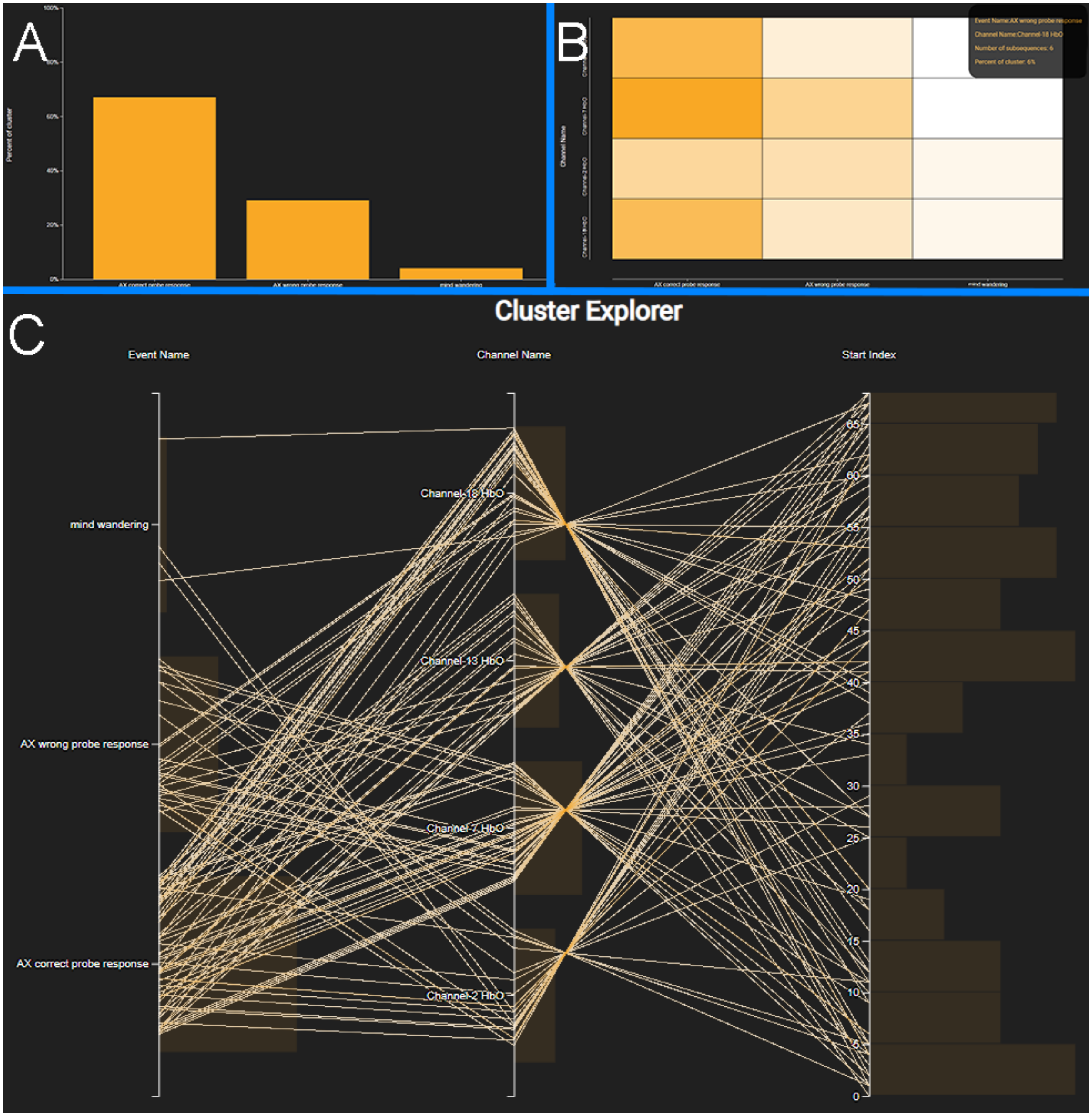


Figure 4.10: The three different vizualizations to explore the content of a cluster. A user will progress from the bar chart (**A**) to the heat map (**B**) then to the parallel coordinates (**C**) as they add additional features to the visualization. **A)** A bar chart displaying the distribution of a single feature in a sample set of sequences. **B)** A heat map displaying the joint distributions of two features in a sample dataset. **C)** A parallel coordinates view displaying the joint distributions between three features in a sample dataset. This visualization will be used for any number of features ≥ 3.

The bar chart has a mouseover component, which allows researchers to explore the precise number and percent prevalence of any value presented.

The second state is a heat map (Figure 4.10.B) that presents the joint distribution over two user-selected features. It uses a linear saturation scale from white to the primary orange or blue color of *BrainEx*, orange when in dark mode and blue when in light mode. White corresponds to 0 percent prevalence, and the full saturation to the highest prevalence present. The use of a consistent, single-hue color scheme means researchers who examine multiple heat maps don't have to learn a new encoding every time. Mouseover text is also available, giving the precise counts and prevalence of the joint distribution.

The third is a parallel coordinates view (Figure 4.10.C). Parallel coordinates allow the visualization of N-dimensional data in 2-dimensional space, making them a powerful choice for researchers who might be interested in any number of data labels [151]. *BrainEx* lets researchers select any number of features, and visualize the joint distribution of all of them at once. Like many implementations, *BrainEx* allows researchers to rearrange the ordering of the axes to explore different pairwise distributions. While increasing the number of features is known to increase the time necessary to explore the data [237], researchers can choose to visualize only features which they are interested in, reducing this exploration time. There are known cases where the display of thousands of data points on a single parallel coordinates map can become unreadable [237]. To accommodate this, we provide a box-select tool, which allows researchers to indicate which sequences they would like highlighted. This allows researchers to separate out a readable subset of the dataset whenever it becomes too large to read.

Once a researcher has some understanding of the overall distribution, they may want to do a more sequence-focused exploration of the data. For this reason, the tool provides a plot sequence feature. This button allows researchers to view a line plot of any set of sequences. This line plot uses color encoding to show the association between the sequences and a single feature of the researcher's selection. This allows researchers to visually check for patterns within the sequence data with respect to the features, or to get a sense of the sort of shapes of series that they are studying. An example of this is in Figure 4.1.

4.5.7 Integration Pipeline

In summary, BrainEx provides researchers the ability to perform similarity searches on time series data. It also provides overviews and visualizations to allow researchers to understand the shape and content of their data. These tools cover the breadth of our first three functional requirements. All of these requirements are important for the exploration of time series data and can be used effectively on their own. However, by integrating the separate tools we can allow researchers to have a more powerful exploration workflow.

When exploring the data distribution within a cluster, researchers may discover interesting time series sequences that they want to explore further outside of the feature distribution exploration. We allow researchers to select a time series from this feature exploration view and then perform a similarity search with that sequence as the target. This allows researchers to better understand the contents of the cluster. Additionally, it allows researchers to find sequences similar to interesting time series discovered in the exploration workflow. This will allow researchers to better understand their data during exploration and potentially reveal previously unknown patterns in the dataset.

In order to integrate *BrainEx*'s exploration capabilities with other systems, our tool also enables researchers to export the contents of similarity search or cluster exploration to a CSV file. This data can then be used in other tools for further analysis after initial exploration and pattern discovery in *BrainEx*.

4.6 Performance Benchmark Summary

To evaluate the scalability and accuracy of *BrainEx*, we conducted a performance benchmark using datasets from the UCR Archive [66], a widely used repository of time-series data. We compared *BrainEx* against three well-established competitors: Piecewise Aggregate Approximation (PAA) [168], Symbolic Aggregate Approximation (SAX) [198], and Generalized Dynamic Time Warping (GDTW) [243] (our ground truth for measuring error, because it provides exact results). These methods represent strong baselines for elastic distance-based time series similarity search.

The experiment assessed both accuracy and response time using three warped distance metrics: warped Euclidean, warped Manhattan, and warped Chebyshev. *BrainEx* consistently outperformed the competitors in both speed and accuracy across small and medium datasets. On average, *BrainEx* was more than 30 times faster on small datasets and over 100 times faster on medium datasets compared to the next best alternatives. Importantly, it maintained error rates below 1% across all distances tested, while the other systems exhibited significantly higher error rates, particularly for Chebyshev distance.

These results demonstrate that *BrainEx* not only supports efficient, interactive querying of large datasets but does so with high accuracy and scalability. For further details on benchmarking methodology and results, we refer readers to the full EICS paper [139].

4.7 Preliminary User Study

To get feedback on the visual exploration interface, we conducted a preliminary user study.

4.7.1 Study Design

For this study, participants were invited to use an instance of *BrainEx* that included a preprocessed dataset containing 8 users with activity in 4 channels. This dataset was generated to represent the usage scenario described in Section 4.5.1. To ensure the study delivered meaningful insights, we sought to draw on the experience of experts in the fields of fNIRS, data visualization, and/or HCI research. The participants were encouraged to spend time exploring the preprocessed dataset in *BrainEx* before answering a questionnaire. The participants were also encouraged to refer back to the application while taking the survey. The study was designed to take approximately 45 minutes to an hour to complete.

Questionnaire

The questionnaire was an online survey developed in Qualtrics. The form consisted of three main sections: one for general demographic questions, the largest section focused on the functional requirements defined in Section 4.5.2, and one section for general *BrainEx* usability and usage questions. The general demographic questions collected background information about the participants' experiences with fNIRS, other brain activity tools, data visualization, and HCI. The functional requirements section of the survey asked study participants to explore a preprocessed dataset through the lens of the usage scenario described in Section 4.5.1. For each of the bullet points within the first four functional requirements, participants were asked to rank how much they agreed with the statement on a 5-point Likert scale. They were also asked to provide any insights they made about the dataset, and any positive or negative comments about their experience completing the requirement. The final section asked the study participants to rate the general usability of *BrainEx* as well as share other comments about *BrainEx*.

Participants

Our study was sent to 40 neuroscience researchers, data visualization experts, and HCI experts who represent our target user base. Of these, 10 responded and participated in our user study. The recruited participants represented a diverse group of target users. The education level and self reported expertise of the participants can be seen in Table 4.1. Six of the participants have published fNIRS or neuroscience research, and five have published HCI papers. Three of the participants had used *BrainEx* before. All participants used Chrome or Firefox to complete the study.

Limitations of the Study

The results of this study are predicated on the subjective responses of the survey participants. We limited the study participants to data visualization researchers, HCI experts, and fNIRS researchers because they are the most likely primary users of *BrainEx*. This is part of what contributed to the small sample size for this preliminary user study. It is also important to note that since the fNIRS research community is small and well-connected, the Likert scale results may experience a positive skew due to familiarity with the research team.

Table 4.1: Participant Demographics

This table provides demographics details for the 10 user study participants. It includes their academic positions and their self-identified expertise in several fields.

Participant	Position	fNIRS	Neural Data Analysis	HCI	Data Visualization
P1	Other	Expert	Knowledgeable	Expert	Knowledgeable
P2	PhD Candidate	Knowledgeable	Passing Knowledge	Knowledgeable	Knowledgeable
P3	Bachelor	Passing Knowledge	No Knowledge	Knowledgeable	Passing Knowledge
P4	PhD Candidate	Knowledgeable	Knowledgeable	Knowledgeable	Knowledgeable
P5	Bachelor	Knowledgeable	Knowledgeable	Passing Knowledge	Knowledgeable
P6	Post Doc	Knowledgeable	Expert	Expert	Knowledgeable
P7	Master	Passing Knowledge	Expert	Expert	Expert
P8	PhD Candidate	Knowledgeable	Passing Knowledge	Knowledgeable	Passing Knowledge
P9	PhD Candidate	Passing Knowledge	Passing Knowledge	Passing Knowledge	Passing Knowledge
P10	PhD Candidate	Knowledgeable	Passing Knowledge	Knowledgeable	Knowledgeable

The dataset used for this study is smaller than most fNIRS datasets and may not be reflective of all possible brain datasets. Thus, we assume some use scenarios may result in future users interacting with *BrainEx* in ways that the study participants did not. The functional requirements defined for this work can be abstracted from our usage scenario to cover possible use cases. In addition, the questionnaire was designed to encourage participants to explore the tool and all of its features.

4.7.2 User Study Results

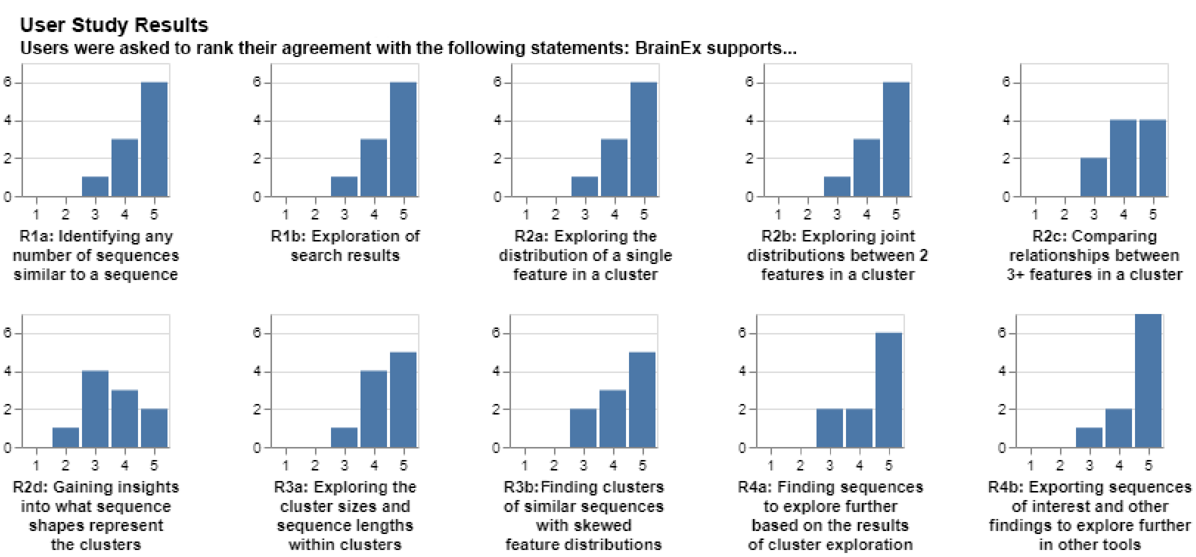


Figure 4.11: The results of the user study. For each plot, the x-axis shows possible responses and the y-axis shows the frequency of each.

Study participants were asked to rank their agreement with statements matching the sub-requirements discussed in Section 4.5.2. The Likert scale covered the range of strongly disagree to strongly agree; these were mapped to the range 1 to 5 for visualization purposes

(Figure 4.11).

Requirement 1: Similarity Search

Participants agreed that *BrainEx* successfully met our functional requirements for similarity search. Participants ranked both sub-requirements very high as seen in Figure 4.11. Nine out of ten participants said they agreed or strongly agreed with the statements, six of the participants strongly agreed with both statements. Six of the participants (P1, P2, P4, P5, P6, P10) provided additional positive feedback in the optional text field informing us the task was very easy to perform and provided results that looked correct and useful. P10 summarized their experience: "*I loved it - very comprehensive. I liked that I could query the most important aspects of the data, and have a fine-grained level of control.*" Despite the positive feedback, P4 and P10 expressed issues with the filter feature of the similarity search.

Requirement 2: Feature Distribution Exploration

Participants were overall satisfied that the system allowed them to explore feature data through *BrainEx*. In general, the more features users attempted to explore, the less strong their agreement. In the case of exploring a single feature, all but one participant (P9) at least somewhat agreed that *BrainEx* supported them, with six participants believing this strongly. The results were similar when considering the visualization of joint distributions. Only one participant (P9) expressed neutrality, and six participants strongly agreed.

The sub-requirements of exploring the distributions of three or more features had weaker, but still favorable results. Two participants (P7, P9) were neutral as to whether the system supported this use case. The rest agreed, but only four participants strongly agreed.

In particular, users expressed an appreciation for the options presented in *BrainEx*. P10 listed the fine level of control as a positive experience when interacting with the system. P5 found the tasks related to feature distribution more challenging than the other tasks, citing the amount of work they had to do in manually examining the data. P6 expressed confusion about the dataset presented in the trial. Despite this, they were able to use the feature-wise visualization to explore the dataset, and make statements about the different user attributes. They felt the software was "*very adaptable*," as they were not bound to specific filters. This suggests *BrainEx* may be useful to analysts who still need to learn more about their dataset of interest.

Requirement 3: Cluster Exploration

Both sub-requirements of the cluster exploration requirement were found to be well supported. Figure 4.11 shows that eight out of ten participants agreed or strongly agreed with requirements 3a and 3b. P5, P6, and P10 liked the fine control provided over what characteristics of the clusters they could view at a time when selecting the cluster. Despite the positive feedback, P2, P4, P5, P6, and P10 found the task of finding an interesting cluster overwhelming due to the number of clusters and desired more features in the cluster exploration tool.

Some of this feedback pertained to features that currently exist in the tool that participants may not have been aware of. For example, the ability to sort and filter the table to more manageable sizes exists as described in Section 4.5.5. Additionally, a toggle is available to show the average feature distribution for the dataset as a baseline for comparison of feature skewness in individual clusters as desired by P10.

Additional features were also suggested in the feedback to improve the efficiency of scanning the cluster list and the ability for cluster comparison. P4 and P10 suggested adding a mechanism for paging through the table quicker and to change the color gradient used for different feature groups to make the cluster table quicker to sift through and easier to separate different feature sets. To improve the ability to compare clusters, P2 suggested being able to view multiple cluster representatives at once so that they could be directly compared.

Future work can make the task of finding an interesting cluster less overwhelming and more informative. Filtering, sorting, and baseline comparison features can be made more exposed to the users and features for more powerful sorting, fine-grained filtering, and navigation through the table can be added. Additionally, the suggested features for comparing multiple clusters and cluster representatives as well as differentiating between feature groups in the table can also be added. However, the results of these survey results still indicate that *BrainEx* successfully supports the exploration of clusters within a dataset.

Requirement 4: Integration

Both of the integration requirements scored very highly. Eight out of ten participants agreed or strongly agree with requirement 4a and nine out of ten participants agreed or strongly agreed with requirement 4b. Requirement 4b was rated as strongly agree by seven participants implying they found *BrainEx*'s ability to export findings into other tools to be very strong. The feedback for this section matched the positivity of the responses. In particular, four participants (P2, P5, P6, P10) filled out an optional text feedback section to share that they thought the integration of the cluster explorer and similarity search was very easy to use and allowed for "*exploring potentially interesting aspects of the data*" (P10). P6 rated this requirement the lowest (neither agreeing or disagreeing with the first sub-requirement), and found the user task somewhat overwhelming and hard to keep track of the patterns and data they were comparing. However, they found the concept of the integration pipeline very promising stating "The pipeline of clustering and search has a lot of potential to explore the data" (P6).

Future work could be done to polish the workflow pipeline and make it easier for analysts to remember the clusters they began the similarity search from. However, the results of this user study show that *BrainEx* successfully completes functional requirement 4 and shows that the cluster exploration and similarity search pipeline provides a novel and powerful exploration workflow for time series exploration.

Requirement 5: Accessible to All Researchers

The accessibility requirement was measured by the ability of participants to use *BrainEx* successfully during the study. Based on feedback of only two users (P2, P8) having bandwidth issues when searching for over a 1000 similar sequences, *BrainEx* supports fast computations. Additionally, based on the positive feedback for the other requirements and the varied expertise and experience of the participants, *BrainEx* can support researchers of all levels. However, improvements can be made to reduce the overwhelming nature of the presented information to further support analysts.

Usability and Additional Feedback

While not surveyed directly, based on the text feedback for each of the functional requirements, participants found *BrainEx* generally usable. Study participants particularly enjoyed the amount of detailed control they had for exploring the data as well as the ease of using the visuals. For example, P1 commented that "*very clear presentation of the results and enable users to explore different attributes*" with regards to the *Similarity Search* requirement. However, participants found the amount of information to be overwhelming at times. For example, P6 noted after the cluster exploration task that "*The number of clusters is often overwhelming*" and that "*It is not easy to identify which ones are most important to look at or to compare a specific small selection of clusters.*" BrainEx does provide filtering on the table of clusters. However, this feedback indicates that this functionality should be better exposed and expanded to allow the user more control over managing the data. While participants did not find severe usability issues with the system, there is room for improvement to expose hidden features, distinguish and clarify elements of the user interface, and allow for better management of large amounts of information.

4.8 fNIRS Case Study

To further illustrate the potential of *BrainEx*, we describe a case study using real-world experiment data from an fNIRS study on cognitive control [140]. This dataset has been analyzed in more traditional methods with mixed effects modeling and investigating the sequence shape. In the study, participants performed the AX-Continuous Performance Task (AX-CPT) which induces different cognitive control states, such as proactive and reactive control [34]. The study concluded that proactive and reactive cognitive control can both be seen in the right dorsolateral prefrontal cortex.

For this case study, we developed a model of the expected fNIRS brain signal when a participant is in each of the two cognitive control states that we were studying. These *model sequences* were based on the expected hemodynamic response, given the tasks and timing of stimuli in the dataset [34]. Using these model signals, we can explore the dataset's clusters to find connections to particular sequences in the dataset.

The goal of this case study was to use *BrainEx*'s cluster exploration to investigate which events and brain regions are associated with the model cognitive control sequences. We identified clusters that had higher-than-average representation of subsequences from the model cognitive control sequences. Our hypothesis was that clusters with a larger number of the model subsequences would contain brain data related to those cognitive states from the AX-CPT task. Therefore, if there was a different distribution of events or brain regions in these clusters compared to the master dataset, then we have identified which events and brain regions are most associated with cognitive control. We expect the cluster results to be different than Howell-Munson et al.'s results because they detected both cognitive control states in the *same* region, while we are looking for clusters that indicate a specific region for each cognitive control state [140].

4.8.1 Dataset Description

The dataset contained the two model cognitive control sequences along with 3,360 sequences from one participant divided between six brain regions. Each sequence is 157 datapoints long and spans approximately 18 seconds of neural data for a total of 527,834 points in the dataset. Along with the brain signal, there is associated metadata for each signal consisting of the *subject name*, *brain region*, *event name*, *start time*, and *end time*. The model sequences have a subject name of "representative" and a region name of "Channel-0" to designate them as different from the collected data. The collected data has six possible region names from the prefrontal cortex (PFC): dorsomedial (DMPFC), left dorsolateral (lDLPFC), right dorsolateral (rDLPFC), ventromedial (VMPFC), left orbitofrontal (lOFPFC), and right orbitofrontal (rOFPFC). There are six possible events with the percentage of the dataset they occupy in parentheses: AX cue (30%), AY cue (20%), BX cue (20%), BY cue (20%), A# cue (5%), and B# cue (5%). These names are associated with the trials in the task, and the details can be found in [34, 140]. We are most interested in AY cue and BX cue, as they can be indications of proactive and reactive cognitive control [34]. Start time is the time in milliseconds when the sequence begins in relation to data collection, and end time is the time in milliseconds when the sequence ended in relation to data collection. The dataset is available here: `https://wp.wpi.edu/hcilab/brainex/`.

We preprocessed the dataset using the Warped Euclidean distance metric, a similarity

Table 4.2: Expected and observed distributions of events from 49 clusters that have high inclusion of modeled cognitive control sequences.

Event	AX	AY	A#	BX	BY	B#
Expected	132,268	88,179	22,044	88,179	88,179	22,044
Observed	127,083	71,795	16,229	92,471	106,561	27,524

threshold of 0.1, and a length of interest of 1-157. In the cluster exploration, we only viewed clusters with a minimum of 20 sequences and a minimum length of 45 (approximately 5 seconds of brain data).

4.8.2 Case Study Results

To ensure the clusters we sampled had a large number of the model cognitive control sequences, we analyzed clusters that were at least 0.15% model sequences, which included 49 clusters. The cluster with the most model sequences had 0.56% which is 11 times greater than the distribution of model sequences in the master dataset. While the ratio of model to participant data is small, this is to be expected as there are 3,780 sequences in the master dataset of participant data and 2 model sequences, making the master dataset consist of 0.05% model sequences.

We used a Chi-square test to determine if the distribution of events and channels in the 49 clusters differed significantly from the distribution in the master dataset. There were a total of 440,895 sequences in the clustered data; the expected and observed distribution of events are located in Table 4.2. Our Chi-square statistic was 10,184, and with 5 degrees of freedom we can reject the null hypothesis ($p < 0.05$) and say the distribution of events in the clusters differs significantly from the master dataset. Notably, all events that start with an A cue appeared less frequently in the clusters than expected, and all events that started with a B cue appeared more frequently in the clusters than expected. The B cue could be indicative of reactive control, one of the modeled cognitive control sequences. Therefore, we can associate these clusters with reactive control for future similarity searches.

The observed and expected distribution for each region in the brain can be found in Table 4.3. Here we also had 5 degrees of freedom and our Chi-square statistic was 887, showing that we can reject the null hypothesis ($p < 0.05$) and say the distribution of brain regions in the clusters differs significantly from the master dataset. Notably, the left hemisphere and VMPFC had a higher frequency in the clustered data while the right hemisphere and DMPFC had a lower frequency in the dataset.

Table 4.3: Expected and observed distributions of channels from 49 clusters that have high inclusion of modeled cognitive control sequences.

Region	DMPFC	lDLPFC	rDLPFC	VMPFC	lOFPFC	rOFPFC
Expected	82,888	62,166	62,166	103,610	62,166	62,166
Observed	80,208	65,808	61,023	110,237	63,139	59,353

4.8.3 Case Study Conclusion

Sometimes a researcher may not know which sequences in a dataset are particularly significant, or which subsequence is a crucial element in their dataset. Through the use of the cluster exploration feature in *BrainEx*, researchers can explore the distribution of the dataset and which sequences are most similar, teasing out meaningful patterns. Through our case study, we demonstrated how one can discover events and brain regions that are correlated with model sequences. The cluster exploration can also help identify which parts of the modeled sequence were most informative by investigating the subsequences that appeared most frequently. For example, in the master dataset, all of the sequences were 157 datapoints long (18 seconds). However, the average sequence length of the clusters was only 80 datapoints (9 seconds) with a maximum sequence length of 126 (14.5 seconds) and minimum sequence length of 45 (5 seconds). By using this information, a researcher can fine-tune their query to be more meaningful to their research questions when using the similarity search feature of *BrainEx*.

4.9 Discussion

Previous research teams have provided basic visual interfaces for DTW based engines to address the increased for interpretability and accessibility by researchers without expertise in using command-line interfaces and APIs (e.g. [242]). These basic visual interfaces for data exploration tools mark a step towards making similarity searches more available [240, 363, 130, 290]; other work focuses on visualizing the results of clustering [129, 189]. *BrainEx* fills a gap in this field by providing a comparison of results across multiple elastic distances and offering insights through combined similarity searches and cluster exploration.

Due to the growing popularity of time series data, there are many other time series exploration tools [108]. *TimeSearcher* provides an interactive similarity search of time series [130]. This tool allows analysts to see a line plot depiction of time series in a dataset. Analysts can use a drag and drop box, known as a *SearchBox*, to select a part of a time series representing an interesting pattern. This pattern can then be queried to discover similar patterns in other sequences. It supports exploration of multivariate data. While it provides a similarity search feature similar to *BrainEx*, *TimeSearcher* does not provide the additional cluster exploration workflow. *QueryLines* is a similarity search tool for time series data that allows analysts to specify soft constraints and preferences which are then used to perform a similarity search on other sequences to discover matching or almost matching patterns [290]. These constraints are added by drawing lines to show the pattern you wish to match. Work by Buono and Simeone into extending the *SearchBox* of *TimeSearcher* showed that drawing a query is an increasingly popular approach in time series similarity searches [38]. Like *TimeSearcher*, *QueryLines* is a similarity search tool and does not provide cluster exploration. *Similan* is a visual similarity search tool for temporal data [363]. It was designed to use a similarity measure called "match and mismatch" to account for temporally misaligned records. The *Similan* researchers propose clustering as a future feature.

Not all time series tools focus on similarity search. Himberg, Hyvarinen, and Espositoc [129] designed a neuroimaging cluster visualization tool using independent component analysis. Kumar et al. proposed a time series clustering tool that represents clusters using a bitmap [189]. These bitmaps can be used for pattern recognition of time series datasets. These tools provide interesting approaches to exploring time series clusters, but do not allow for exploration via similarity search.

4.10 Future Work

Future work on the *BrainEx* engine could focus on making the preprocessing step of *BrainEx* even more efficient through converting the code into another language, such as Rust. Additionally, *BrainEx* could replicate studies with fNIRS curated specifically for validating tools, such as the n-back dataset from Wang et al. [355]. While we used the original versions of SAX and PAA to do our benchmark comparison with *BrainEx*, newer versions of these algorithms exist and can be tested against *BrainEx* [318, 378]. These versions were out of the scope of the experiment presented in Section 4.6.

In addition, further refinement and evaluation of the interactive visual interface could improve the user experience. Future user studies could aim to provide the participants with a larger fNIRS dataset to explore with the interface. In addition, a larger cohort of participants could be recruited from a more diverse set of expertise to be able to investigate the differences in the usability of the tool between experts in data visualization and neuroscience compared to researchers who are just starting their scientific careers.

To promote collaborative research and accessibility, we created a website (`https://wp.wpi.edu/hcilab/brainex/` to make the *BrainEx* code available to researchers. In addition, the results from the performance experiment and the clusters from the case study can be found there.

4.11 Conclusion

We present *BrainEx* as a tool for visual exploration of brain signals. By combining cluster exploration, feature distribution exploration, and similarity search we provide a powerful and novel exploration workflow that existing fNIRS and time series analysis tools do not provide. In our performance experiment, we demonstrated how *BrainEx* is lightning fast compared to state of the art competitors as well as highly accurate. We developed five functional user requirements, and based on the results of a preliminary user study with HCI and neuroscience researchers, we determined *BrainEx* meets these requirements. Finally, we used a case study to demonstrate how a researcher could use *BrainEx* to make inferences about real-world fNIRS brain data. Overall, we have shown that *BrainEx* could be an effective tool for fNIRS or other neuroscience researchers.

Chapter 5

BCIs for Teamwork: Relevant Background

Here we present relevant literature and prior work across the domains of social psychology, human-computer interaction, and neuroergonomics to provide a foundation for our studies. First we discuss the nature of teams and creative collaboration, how creative synergy can arise, and pitfalls that can result in process loss. Then we briefly discuss the existing work on creativity support tools and BCIs for creative applications and how BCIs can be used to support creative collaboration. Finally, we summarize prior studies that assessed perceptions of BCIs and other emerging technologies.

5.1 Teams and Creativity

In sociology and social psychology, a group is typically defined as a collection of two or more individuals connected by and within social relationships [97]. Likewise, a team is a particular type of group whose members work together in pursuit of a shared goal [97]. Teams work together as an organizational unit—failure or success occurs at the group level, rather than for members individually. Beyond this, however, teams can vary considerably in size, composition, or structure [260], with global trends such as digitalization and globalization increasingly challenging the traditional notion of what teams are and how they function [349]. Team composition can be selected by members (e.g., for a school project), or by organizational leadership (e.g., for a company division); members can be co-located, or dispersed to various degrees technologically and geographically [120]; and have varying distributions of skills, knowledge, and expertise [349]. The bounded and stable membership of classical teams has frequently given way to teams in which membership changes frequently, and members taking on multiple roles, or even roles in multiple interdependent teams. Organizations have in some instances encouraged the formation of self-organizing and self-governing teams, in addition to (or in lieu of) traditional top-down leadership structures.

As illustrated above, teams are complex dynamical systems, and their success is a function of the social and environmental context in which they are operating as well as team composition (demographics, cultural background, expertise, leadership and pacing styles, etc.) [79, 23]. These in turn give rise to affective states, cognitive states, and behavioral processes, which can evolve over time as team composition changes, conflicts occur, or trust grows between team members. Such states are emergent both at the individual level (e.g., the valence, arousal, cognitive load, and trust of individual team members) and at the team level (in the case of team processes such as coordination, strategy formulation, affect management, etc.) [30, 218]. Work by Woolley et al. [364] has shown that the contribution of these processes toward task performance can be summarized by a collective intelligence (CI) factor, which is predictive of success across a variety of group tasks [287, 170]. Teams with high CI tend to have members with greater social perceptiveness, and a greater proportion of women. In general, gender and ethnic diversity is associated with greater CI in teams [54]; however, the level of group and individual processes of team members is a greater predictor of CI than team composition [287]. Interpersonal emotion regulation (IER), when individuals engage in actions motivated to modify emotional expressions or experiences, may also impact the quality of relationships between team members. IER strategies that are appropriate for the given context may positively impact the affect, well-being, and interpersonal closeness of team members [73], with engagement-oriented strategies, such as resolving conflicts and seeking emotional support, predictive of lower feelings of loneliness and greater feelings of connection and relationship satisfaction [256].

Creativity is defined by Paulus, Dzindolet, & Kohn [260] as "the generation or production of novel products or ideas." While much of the early research on creativity focused on creativity of individuals [191, 119, 313, 314], more recent work has investigated creativity at the team level. Teams engaging in creative tasks are in many respects equivalent to other teams, and the goal of collaboration broadly remains the same: that of realizing synergy, whereby team performance is enhanced relative to that of individuals. Thus, supporting teamwork in general can be advantageous to team creativity as well. However, there are several factors that can have an outsized influence on a team's creative potential. One model of high-performing creative teams (e.g., interdisciplinary teams at Pixar) suggests that teams formed from subgroups from different disconnected disciplines (e.g., art and technology) with different skillsets and expertise, rather than merely just a large array of diverse viewpoints and perspectives, are most successful [125]. Thus, a team that has a membership with diverse perspectives (experiences, expertise, problem-solving strategies) that are task-relevant can draw on them to explore a greater number of ideas, provided the work environment facilitates integrating and exchanging these perspectives [261, 259]. Fostering a work environment where team members feel psychologically safe enough to share their ideas without fear of discouraging and harsh feedback, while also allowing some amount of conflict and constructive criticism, is likewise necessary for maximizing creative potential

[191]. Furthermore, the generation of novel ideas relies upon lateral or associative thinking (in which thoughts leap from category to category via semantic associations) and divergent thinking (i.e., free-flowing, non-linear thinking where ideas are generated in an emergent fashion) [353, 325, 357, 182]. The degree to which team members can successfully employ lateral and divergent modes of thinking is dependent on the dynamic interplay of cognitive states and processes active during ideation [47, 342]. Cognitive flexibility (the ability to update goals and actions in response to changing contexts and task demands) may also play a role [78, 135].

Despite their potential for creative synergy, there are several challenges creative teams face that can lead to process loss. Behaviors such as social loafing (when some group members put forth less due to the presumption that others will contribute more) or social comparisons with other team members, as well as lack of internal or external motivation (e.g., due to lack of trust between team members or incompatibilities between team members and leaders) hinder creative output [191, 223, 55]. Verbally expressing ideas can lead to production blocking, where participants forget their ideas while waiting for a turn to speak, or choose not to share them because they feel they are less relevant after time has passed, and eventually become discouraged from sharing future ideas [248]. Alternatively, rehearsing ideas internally can prevent group members from being attentive to the ideas of others. While asynchronous remote collaboration has the potential to alleviate these issues by allowing team members to record their ideas in real time [261], remote collaboration has been shown to curb idea generation by narrowing the cognitive focus of communicators to screens [36]. The potential for remote collaboration technology to be misused for surveillance, coupled with the always-on nature of digital communication channels, can also impede collaboration by reducing the time available for cognitive and social processes required for divergent thinking and the integration of others' perspectives [154]. The consensus in the literature seems to be that teams need balance among several qualities to avoid process loss and maximize synergy [260]. Teams should have some amount of cohesion and trust, but not so much that interpersonal bonds between team members are prioritized over commitment to the task at hand. Teams should have some amount of diversity in task-relevant expertise and experiences, but teams that are too diverse may have difficulty integrating knowledge across domains. Team members require both intrinsic motivation and external support (e.g., a work environment that facilitates creative thinking and collaboration, where members feel safe to share their ideas [347, 226], as well as tasks and goals set by team leaders). Teams can also benefit from workflows that integrate remote and in-person collaboration, to mitigate the pitfalls of each. These approaches to enhance creative synergy ultimately aim to benefit the affective states, cognitive states, and behavioral processes that belie collaboration.

5.2 Artistic Brain-Computer Interfaces and Creativity Support Tools

Creative professionals have historically been quick to experiment with and adopt novel technologies to augment art and ideation [83]. From recording devices such as the phonograph or camera, to graphic design software and generative artificial intelligence programs, technology has shaped the ways artists are able to create emotionally engaging and thought-provoking experiences for audiences. Artists have been using physiological sensing and brain-computer interfaces to shape such experiences for decades. The first use of a BCI for artistic exploration was in 1934, where alpha waves recorded via EEG were converted to sound. American composer Alvin Lucier later used the technique in concert in 1965 [272]. Since then, BCIs and recordings of physiological signals have—via passive or active control by users—been used to create music [235], generate lighting/sound effects during performances in response to audience engagement [369], alter the narrative flow of movies [266, 283], provide visual or auditory representations of users' cognitive or affective states, play games [52], and more [9].

In addition to being integrated with art exhibits and experiences directly, researchers have also demonstrated how BCIs can serve as creativity support tools, which are digital systems that can enhance creativity by assisting users of varying levels of expertise in one or more phases of the creative process (e.g., planning, ideation, implementation, evaluation, and iteration) [101]. A wide variety of such tools have been developed both for individuals [354] and groups [102], including tools which help define problem scope of prior to brainstorming [22], allow users to map the semantic connections between ideas [303], iteratively generate new graphics based on initial user input [145], and more [354]. BCI-based support tools are additionally able to respond to the cognitive and affective states of users. Botrel, Holz, & Kübler [32] and Todd et al. [328] developed hands-free, brain-powered graphic design tools which used the P300 event-related potential, a response to conscious decision making. Other tools include an artificial agent that provides design suggestions for architectural designers based on their affective state [371], neurocognitive feedback to enhance creative problem solving [141], and a brainstorming assistant that can alter the semantic distance between suggested ideas in response to a user's level of cognitive effort [47]. However, while these tools are able to leverage brain activity to provide novel forms of assistance, thus far the BCI-based tools in the literature collect data from and provide assistance to individual users. No existing BCI-based creative support tools have simultaneously recorded and processed the brain data of multiple users to assist with creative collaboration.

5.3 Brain-Computer Interfaces to Support Teamwork

Given that the affective states, cognitive states, and behavioral processes of team members are indicative (and sometimes predictive ([287, 170]) of successful collaboration, measuring these states and processes in real-time as teams are working could provide useful input to a tool or interface designed to aid teams in achieving synergy during collaboration. While prior work has demonstrated that affective states, such as valence and arousal [143, 84, 65], and cognitive states such as mind wandering [74], multitasking [307, 5], learning stages [138], and workload [277, 13, 311] can be measured and integrated into online systems via continuous non-invasive physiological sensing, determining the presence and magnitude of team processes has generally relied on administering periodic surveys and behavioral assessments [218, 225], which can disrupt team workflow and only provide sporadic indicators of how well they are collaborating. However, some work employing hyperscanning (simultaneous recording) using electroencephalography (EEG) [75] and fNIRS [64, 138] has demonstrated differences in neurophysiological signals between participants working individually versus as a team, and between expert and novice teams.

Eloy and colleagues [85] demonstrated that fNIRS can be used to measure levels of team processes in real-time. Similar to functional magnetic resonance imaging (fMRI), fNIRS measures hemodynamic responses that occur in the brain following neuronal activation. Instead of relying on an expensive and non-portable MRI scanner, fNIRS employs emitters and detectors of near-infrared light placed on the scalp to determine the concentrations of oxygenated and deoxygenated hemoglobin in the brain. fNIRS has temporal resolution that is comparable to fMRI and boasts greater spatial resolution than electroencephalography (EEG) [1]. Additionally, fNIRS devices are low-cost, portable, and require little advance preparation, making them a good choice for real-world scenarios [309]. By employing fNIRS hyperscanning (simultaneous recording) while pairs of users collaborated with an artificial agent during a realistic resource allocation task, Eloy and colleagues were able to reliably model the levels of different team processes (coordination, strategy formation, and affect management).

5.4 Public Perceptions of BCIs and Emerging Technologies

As a BCI support system for creative collaboration does not currently exist, we look to existing work examining perceptions of BCIs and other emerging technologies to inform the design of our study exploring stakeholder needs (Chapter 7). Studies generally choose either of two main approaches: conducting surveys of large samples representative of the general public or demographics of interest, or semi-structured interviews of smaller, targeted groups of participants (usually with a design probe to stimulate thinking and discussion).

An example of a relevant larger-scale study is work by Sample et al. [291], who conducted a large ($N = 1,403$) web-based transnational survey of the public in Germany, Canada, and Spain assessing attitudes toward ethical issues related to BCIs. They found participants had moderate concern for agent-related issues (changing self-perception, stigma, autonomy) and consequence-related issues (new forms of hacking, privacy concerns, moral/legal accountability), with women and those who were religious slightly more likely to have stronger concerns. A related study by Sattler & Pietralla [293] used a nationally representative sample of German adults ($N = 1,089$) asking about their willingness to use BCIs and whether they were morally acceptable. In a $2 \times 2 \times 2$ factorial experiment, the authors varied purpose (treatment vs. enhancement), invasiveness (invasive vs. non-invasive) and framing (i.e., the order questions assessing moral acceptability vs. willingness were presented) in vignettes that introduced BCI technology to participants, and moderate moral acceptability and willingness to use, with a preference for treatment over enhancement use cases and non-invasive over invasive devices. Finally, Tindale et al. [326] surveyed 344 people (employers and employees) about the use of brain and body signals in the workplace across a wide variety of occupations (construction, healthcare, government, education, etc.) in British Columbia, Canada. While 95% of participants did not use brain or body signals in their workplace, benefits for it they envisioned included uses for health monitoring, wellness, and safety, while potential risks included stress, a lack of privacy, and excessive oversight of employees by employers. Interestingly, brain sensors were more likely to be seen as inconvenient and having no benefit compared to body sensors, and a majority of participants said employees should own the info recorded from them (with employers more likely than employees themselves to say this). However, while these larger studies can gauge the perspectives of the public and get a clearer picture of trends for specific sub-populations, these studies miss the chance to ask probing questions to further explore perspectives of individual participants and could miss out on unexpected insight.

In contrast with the larger studies above, studies that recruited smaller, more targeted cohorts of participants were better able to engage with particular groups of interest in greater depth. Merrill & Chuang [230] wanted to assess software developers' narratives and anxieties around BCIs and explore their visions of the future for these devices—how do their ideas about the mind and how it relates to the brain and body inform and constrain beliefs about what BCIs can and should do? The authors conducted semi-structured interviews with 8 Silicon Valley software engineers who did not have prior experience working with BCIs, using a user authentication BCI built with a Muse EEG headset as a design probe to inspire thoughts and questions from interviewees. Participants voiced different positions on what the mind was (e.g., the brain acting like a computer; "conscious awareness" separate from unconscious phenomena that affect the mind; embodied cognition arising from the connection between brain, body, and the environment), but all believed BCIs could "read" and decode it. The probe caused participants to speculate on the future of BCI technology generally. Similar to the Tindale et al. study above, the chief concerns participants had related to privacy and security, with several wary of BCIs being able to "leak their thoughts." The consensus among participants was that BCIs would become pervasive (one said we would have to "come to terms with [them]") whether we wanted them to or not, rather than individuals having agency about using them.

Devendorf et al. [71] were interested in examining the future of technology and fashion, and explored an emerging technology altogether different from BCIs: computationally responsive clothing. The authors used Ebb, a textile display made from conductive thread coated with thermochromatic paint, as a design probe in semi-structured interviews where 12 potential wearers and 5 designers compare it to existing similar technologies (e.g., fiber optic or flexible LED screens). The interviews elicited several novel use cases in real-world contexts—Ebb could be used to connect physical and digital lives, display physiological data, show time or transit info, have many-in-one garments, and more.

Merrill & Chuang [230] and Devendorf et al. [71] both used design probes in their studies, but what if a prototype is not available, or has not adequately considered the needs of users? Holstein, McLaren, & Aleven [133] conducted interviews with 10 middle school teachers to explore their needs for intelligent tutoring systems (ITSs) with the goal of designing a real-time ITS. Instead of using a design probe to facilitate this exploration, the researchers relied on a combination of other techniques, so the assessment of needs would inform the development of a prototype, and not the other way around. First, the researchers conducted design interviews using superpowers as a probe, asking the teachers, "If you could have any superpowers you wanted, to help you do your job, what would they be?" The teachers wrote responses on index cards and were asked to prioritize them. This allowed the researchers to get a sense for teachers' needs in the classroom, where breakdowns in current practices occurred, without feeling constrained by solutions offered by existing technology. To inquire about the teachers' needs more directly, the researchers conducted semi-structured interviews asking about their experiences with ITSs, and how they could be improved. Many teachers were concerned that ITSs were *replacing* their roles rather than *supporting* them, and desired analytics that could provide more valuable insights about either their teaching or their students learning. Finally, participants "speed dated" several possible futures based on their superpowers depicted with storyboards in quick succession. Researchers used this to probe boundaries of what participants considered acceptable system behavior, and discovered that while teachers appreciated designs that presented them with information that could help them prioritize their time, they disliked alert systems with direct recommendations because they were perceived as threatening their autonomy in the classroom.

In summary, while their results may not generalize to larger populations, the studies with smaller cohorts above demonstrate several advantages over studies with larger pools of participants. Smaller studies allow researchers to target specific populations more easily, and it is easier to try a variety of different approaches with the same participants or ask probing questions to further explore ideas that arise in discussion. Additionally, meeting with participants face-to-face allows researchers to present tangible artifacts to participants that they can interact with, which can elicit perspectives that may be difficult to attain otherwise. All these advantages make adopting the methodological approaches they used appealing for our present studies.

Chapter 6

BCIs for Teamwork: Exploring the Relationships Between Brain Markers for Individual and Group States During Collaboration

Abstract

This exploratory study investigates the relationships between individual- and group-level neural measures, team behavior, and collaborative performance in small groups engaged in a creative design task. Participants collaborated to design a virtual escape room using digital tools and an AI assistant, while EEG recordings captured both individual cognitive states and inter-brain synchrony dynamics. Analyses revealed modest but meaningful associations between neural activity and behavioral dimensions of collaboration, aligning with prior research linking group neural dynamics to team processes. These findings demonstrate the potential for brain-computer interface (BCI) systems to leverage such signals in supporting teams during collaboration.

6.1 Introduction

Creativity is key for developing novel art and ideas, tackling complex challenges, and generally driving innovation. Several convergent factors—individuals' backgrounds, behaviors, and personalities; their cognitive processes; and the environment and broader social context in which they operate—influence creative thinking and output [191]. Thus, properly accounting for and manipulating these variables is necessary in order to maximize creative production [28, 226]. Additionally, prior work has shown that collaboration between individuals has the potential to fuel creative synergy, where new cognitive inputs, the combinations of personality characteristics, or interaction dynamics can yield a volume, breadth, and fluency of creativity greater than what would be possible from working alone [191].

However, an increase in creative output as a result of collaboration is not guaranteed; indeed, several studies have shown that creative collaboration can result in *decreased* output, despite the potential gains. Such process loss can be caused by a number of mechanisms, including groupthink (the tendency of group members to adopt the majority perspective), social anxiety or apprehension toward sharing ideas, downward comparison with others, cognitive load, or distraction by other group members [191, 85, 68, 258, 261]. Thus, while it is possible to have high-performing creative teams, the appropriate environment, resources, and workflow must be fostered for creative synergy to occur.

Monitoring and supporting effective cognitive states and team processes is critical for enabling teams to reach their full potential while minimizing process loss. Adaptive support systems that can operate in real-time and unobtrusively—without disrupting ongoing interaction—may help facilitate productive social and cognitive activities such as coordination, negotiation, and planning [325, 260]. Recent advances in neuroergonomics suggest that such systems could benefit from the integration of physiological and neurophysiological sensing modalities, including EEG, which has been shown to capture dynamic changes in cognitive states like attention, workload, and mind-wandering with relatively high temporal resolution [381, 197]. For example, EEG-based indices have successfully been used to detect lapses in attention and cognitive fatigue, supporting their relevance in collaborative scenarios [197].

In addition to individual-level metrics, surface brain sensing approaches such as EEG and functional near-infrared spectroscopy (fNIRS) have also been used to investigate interpersonal neural synchrony and coordination dynamics in social and collaborative contexts [85, 64]. Various measures including wavelet transform coherence [64], the recurrence rate from multidimensional recurrence quantification analysis [85], mutual information [315], and others have been demonstrated to assess joint mental processes during collaboration.

In this chapter, we report on an exploratory study examining teams as they perform a creative collaboration task (designing a virtual escape room with the help of an AI assistant) while undergoing brain recording via EEG. We investigate how both individual measures of cognitive engagement and workload and group measures of brain dynamics relate to team behaviors and performance, as well as how individual-level neural dynamics correspond with group-level synchrony.

We ask the following research questions:

1. What is the relationship between neural measures of individual, internal states (relative band power, engagement, workload) and neural measures of joint interpersonal states (measures of inter-brain synchrony)?

2. What is the relationship between group performance and behavior, individual emotion regulation strategies, and brain state measures?

Our aim is to gain a better understanding of the relationships between internal states and interpersonal coordination in real-world team contexts, with the goal of informing the future development of neuroadaptive systems to support team collaboration.

6.2 Study Design

This exploratory study was a multi-institute collaboration between Worcester Polytechnic Institute (WPI) and the University of Bremen, which leveraged behavioral and physiological data collected from team members working together to complete a complex creative task (designing a virtual escape room) in order to gain a more comprehensive understanding of collaboration dynamics in realistic scenarios. To capture the complexity of real-world work teams and the current roles technologies can play in facilitating synchronous collaboration, team members were distributed across both sites and collaborated remotely, with access to a digital whiteboard and a generative AI assistant.

6.2.1 Participants

Participants were 44 members of the WPI and University of Bremen communities with no known neurological conditions (median age range 18–24, 26 male; full demographics in Table B.1) grouped into 12 teams of three or four members each (max two people per institution) collaborating remotely. To help ensure that we could capture a variety of team dynamics over the course of the study, we did not require that participants knew each other beforehand, nor did we require that they were familiar with escape rooms or generative AI tools (they were briefed on these prior to the study). We anticipated that varying degrees of familiarity with their teammates, the task, and AI tools would lead to a spectrum of more and less successful teams, as well as different levels of the variables of interest (e.g., workload, team coordination). Participants were compensated $15/hr or €15/hr, depending on their respective locations, for their time. Recruitment and experimental procedures were approved both by WPI's Institutional Review Board and the University of Bremen's Ethics Committee. Participants completed informed consent forms upon arriving at each lab.

6.2.2 Data Acquisition and Experiment Procedure

Participants worked in teams of three or four (up to two from each respective institution) to design a virtual escape room—an immersive game where one or more players work together to solve puzzles and complete challenges to "escape" within a set time. In anticipation of their widespread use by teams in the future, we also allowed participants to employ the assistance of ChatGPT, a generative AI chat assistant. This design task was chosen because it represents a complex, realistic creative problem-solving scenario that lends itself to multi-person collaboration (requiring the exchange of information and the establishment of a joint mental model), while not being so straightforward that assistance from an AI tool would trivialize it. Furthermore, prior work has shown that engaging in such creative collaboration tasks can elicit group-level physiological phenomena such as interbrain synchrony [64, 207, 366], which could serve as real-time indicators of the quality of team interactions.

Participation took place remotely over Zoom to facilitate video and audio recording. Participants were still physically present in separate rooms at their respective institutions so physiological data could be collected. Participants had one hour to work together to design a virtual escape room with help available from ChatGPT, and were required to document their design via a shared Miro digital whiteboard. The long duration of the experiment helped ensure that states observed in real-world teams such as distraction, disengagement, or internal reflection naturally occurred.

Prior to the experiment session, participants were provided an overview of the study goals and experiment procedure and a copy of the informed consent form. In order to ensure a common baseline understanding, we also provided a briefing document with information about what virtual escape rooms are, important elements to include in their design, and how ChatGPT could be used to assist them. Using Krekhov et al.'s [184] taxonomy for escape room games as a foundation, we described several items and questions participants should consider, including

- **Mental Challenges:** Puzzles involving observation, pattern recognition, calculation, and knowledge.
- **Physical Challenges:** Tasks requiring object movement, alignment, or agility.
- **Emotional Challenges:** Elements that evoke strong emotions (e.g., unease, fear, surprise), require difficult decisions, or deal with negative consequences.
- **How many players** is the challenge made for?
- **Why** are players trapped in a room? What is the **theme or narrative** for the challenge?
- Will players be able to **receive hints** if they are stuck?
- **How much time** do players have to escape?

Finally, we provided a copy of the rubric that would be used to score their escape room design after the session, likewise based on the taxonomy from [184]; the rubric criteria and scoring scheme are provided in Table 6.1.

We also administered a brief pre-experiment survey via Qualtrics to collect demographics data, measure participants' familiarity with escape rooms (1 - Not at all familiar (I've never heard of them before); 2 - Somewhat familiar (I know about them, but I've never tried one myself); 3 - Moderately familiar (I've been to physical escape rooms or played virtual escape room games once or twice); 4 - Very familiar (I have been to several physical escape rooms and/or played several virtual escape room games, or have designed escape room experiences)), and assess differences in their emotion regulation strategies using the Process Model of Emotion Regulation Questionnaire (PMERQ) [256].

During the experiment session, participants were first introduced to the study goals and experiment procedure, and given the opportunity to ask the researchers any outstanding questions. If after the briefing they still wished to participate, participants signed a physical version of the consent form and the experiment session proceeded. Participants were advised that they could withdraw from the study at any time.

Participants were each seated in front of a laptop or desktop computer at their respective locations, all in separate rooms to ensure audio isolation. On each computer, participants had access to a common Zoom meeting allowing for video communication with the researchers and other participants, as well as the Chrome browser with tabs open to the shared Miro digital whiteboard, ChatGPT, and a copy of the rubric that would be used to score designs. Each participant's camera feed and screen, along with microphone and system audio, were recorded simultaneously using Open Broadcaster Software (OBS) Studio [18]. To ensure the audio and video could be aligned with recorded brain data, we also integrated an OBS plugin[1] which allowed relevant timing information to be published to a data stream in Lab Streaming Layer (LSL) [180], an open source middleware framework for streaming, receiving, and synchronizing multimodal time series. Specifically, the LSL plugin generates a filter overlaying the recorded video containing the frame number and UNIX timestamp as measured by the clock of the recording computer, and publishes an LSL stream with this information that can be recorded by one or more other computers on the same local network.

Additionally, participants' brain activity was recorded via EEG at both study locations using g.tec Unicorn Hybrid Black headsets (g.tec medical engineering GmbH, Scheidlberg, Austria [114]). Each headset recorded electrical activity from eight rubber electrodes across the cortex (located at Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8 according to the International 10-10 system [252]) at 250 Hz, using two mastoid references. Conductive gel was applied to each electrode site to ensure optimal signal quality, which was assessed with the Bandpower component of the g.tec Unicorn Suite software prior to recording. Participants were instructed to silence their cell phones and place them on the table away from any equipment for the duration of the experiment to minimize electrical interference. Once the signal for each channel had stabilized and acceptable quality was confirmed, a custom Python package utilizing the `pylsl` library[2] was used to publish LSL streams for each Unicorn device containing their respective channel data.

[1]https://gitlab.unige.ch/sims/lsl-modules/obs-plugin
[2]https://pypi.org/project/pylsl/

Category	Criteria	Description	Points
Puzzle Variety (30 pts)	Mental Challenges	Range and balance of observation, pattern recognition, calculation, and knowledge puzzles.	**10 pts** Excellent **7 pts** Good **4 pts** Average **1 pt** Poor **0 pts** Missing
	Physical Challenges	Inclusion of object movement, alignment, agility, and timing tasks.	
	Emotional Challenges	Incorporation of puzzles that evoke and require overcoming emotions.	
Thematic Coherence (20 pts)	Integration with Puzzles	How well the theme is integrated with the puzzles and overall design.	
	Consistency and Creativity	Consistency and creative use of the theme throughout the escape room.	
Narrative Integration (20 pts)	Story Depth	Depth and engagement of the storyline.	
	Story-driven Puzzles	How well the narrative drives the puzzle-solving process.	
Technical Feasibility (15 pts)	Practicality of Implementation	Practicality and feasibility of implementing the design in a virtual environment.	
	Effective Use of Technology	Effective use of available technology without complicating the user experience.	**5 pts** Excellent **3.5 pts** Good **2 pts** Average **0.5 pts** Poor **0 pts** Missing
User Experience (10 pts)	Intuitiveness and Accessibility	Ease of navigation, clarity of instructions, and overall user-friendliness.	
	Engagement and Enjoyment	How engaging and enjoyable the experience is for players.	
Innovation (5 pts)	Creativity in Puzzle Mechanics	Creativity and uniqueness in puzzle design and room features.	

Table 6.1: Scoring rubric used for escape room designs, based on the taxonomy of digital escape rooms by Krekhov et al. [184]. A maximum of ten points could be earned for each of the first eight criteria, an a maximum of four points could be earned for each of the last four, for a total possible score of 100 points.

Once the OBS and Unicorn LSL streams for each participant were published, the streams were recorded on at least one computer at each site using the LabRecorder application[3]. All computers were connected to a common virtual private network (VPN) using the ZeroTier VPN client[4] to ensure that all streams were visible to the computers at both sites. Video and audio recording through OBS was started a few seconds after beginning to record LSL streams in LabRecorder, due to instability in OBS with simultaneous remote-controlled start.

Finally, after the data recording setup was complete, participants were ready to begin the study. Participants were first provided a brief overview of the Miro interface, and then given directions for the design task. Participants were told that they had one hour to design a virtual escape room according to the criteria in the rubric, and that they should aim to score as high as possible, though given the time constraints it was understood that designs would not be perfect. We advised participants that they did ***not*** need to build or implement their designs; all that was required was that they work together to come up with a design and document their design process using Miro. We also advised that participants might prefer to delegate responsibilities for using each of the available tools, or divide their use equally, and that because data provided to the tools or shown to others was not necessarily protected, that participants should refrain from entering or verbalizing personal information they wanted to omit from the recorded data.

After receiving these directions, participants were instructed to begin the task, and a one-hour timer was started. Researchers remained in the Zoom call with participants to answer any questions, but turned off their audio and video feeds to avoid distraction. When the timer elapsed, participants were instructed to add any final touches to their designs and then cease all work, after which compensation was provided.

6.3 Analysis Methods

Due to the rich multimodal nature of our dataset, we employed several analysis techniques to explore the different levels of our data and their relationships with one another. Here we describe the preprocessing and analysis methods we used to analyze group performance, participant emotion regulation strategies, their behavior during the task, the recorded EEG data on both the individual and group levels, and the relationships between these different measures.

6.3.1 Task Performance

Task performance was scored using the rubric in Table 6.1. Two researchers first individually scored each group's designs as documented on the Miro board (referring to the session videos

[3]https://github.com/labstreaminglayer/App-LabRecorder
[4]https://www.zerotier.com/

if necessary for clarification), providing numeric scores and notes explaining the reasoning for the score choice for each of the 12 design criteria. Then, if the researchers' scores for any criteria differed by more than one ranking (i.e., if one researcher rated one item as "Excellent" and the other rated it as "Average," or if one rated an item as "Good" where the other rated it as "Poor"), the researchers discussed their reasoning for assigning their respective scores and deliberated until the discrepancy was resolved, and scores differed by at most one item ranking. Finally, the final score was calculated by summing the average scores for each criterion. (E.g., if Researcher 1 assigned a score of 10 for "Mental Challenges" and Researcher 2 assigned a score of 7, the final score for that criterion was 8.5.)

6.3.2 Emotion Regulation Strategies

Emotion regulation strategies for each participant were assessed prior to the experiment using the Process Model for Emotion Regulation Questionnaire (PMERQ) [256]. The PMERQ assesses the degree to which respondents' emotion regulation strategies are either engagement-focused or disengagement-focused in each of the five stages of emotion generation as described in the *process model of emotion regulation*. These stages occur in sequence in response to an emotion-generating situation; at each, one can intervene to change the emotion, typically to increase positive emotion or decrease negative emotion.

According to Olderbak et al. [256], the first stage, *situation selection*, involves taking action in anticipation of a situation, such as by confronting or avoiding it. The second, *situation modification*, involves taking action in response to the situation, such as resolving or sidestepping conflicts. The third, *attentional deployment*, involves changing one's attention to the situation, such as directing attention away from certain information and toward other information, (e.g., away from people yelling and toward a supportive friend or authority figure). The fourth, *cognitive change*, involves reframing how one thinks about the current situation to alter how one feels about it. Finally, *response modulation* occurs after the emotion has been generated and involves changing components of the generated emotion such as subjective feelings, physiological arousal, and expressive behaviors.

At each stage, an individual's regulation strategies can either be engagement-focused (e.g., confronting/facing the situation) or disengagement-focused (e.g., avoiding or removing oneself from the situation); such strategies are thought to be *trait-like*, and stable for individuals across different situations. In general, engagement-focused strategies are associated with higher well-being and better relationship functioning than those which are disengagement-focused.

The PMERQ consists of 45 questions to assess the degree of engagement- or disengagement-focus of an individual's emotion regulation strategies during each of the five stages, resulting in subscales for 10 strategies in total (2 focus types $\times$ 5 stages); see Table 6.2. Each item was a statement relating to one of the 10 emotion regulation strategies (e.g., "I work to

Emotion Regulation Stage	Engagement-Focus	Disengagement-Focus
Situation Selection	Confront Unpleasant Situations 4 items	Avoid Unpleasant Situations 6 items
Situation Modification	Resolve Conflicts 6 items	Avoid Conflicts 5 items
Attentional Deployment	Focus Elsewhere 4 items	Cognitively Distract 5 items
Cognitive Reappraisal	Consider Benefits 6 items	Reduce Importance 3 items
Response Modulation	Support by Emotion Sharing 3 items	Expressive Suppression 3 items

Table 6.2: Emotion regulation strategies measured by the PMERQ, and the number of associated survey items. The final score for each strategy is the average of item-level responses.

negotiate a resolution to conflicts I have with others, to decrease how bad I feel."), to which participants indicated their level of agreement from 1 (Strongly Disagree) to 6 (Strongly Agree). The final score for each strategy was the mean of its item-level responses. Three participants failed to complete the pre-experiment survey, leaving 41 complete responses.

6.3.3 Team Processes from Speech and Behavior

In order to gain meaningful insights into the processes and dynamics at play during teammate interactions and facilitate comparison with physiological data, we employed approaches from interaction analysis [33] to analyze the audio and video of participants during their experiment sessions. To quantify the levels of team processes present during verbal interactions between team members, we adapted the approach from [221], who developed and validated a dictionary for performing computer-aided text analysis (CATA) to measure the 10 dimensions of Marks et al.'s [218] taxonomy of team processes, which has been used widely in prior work [33, 85, 171, 221, 289, 361]. Briefly, Marks et al.'s framework describes a *recurring phase model* of team interactions, in which teams perform in temporal cycles of goal-driven activity marked by identifiable periods of action and transition between actions, with the management of interpersonal relationships occurring throughout. These episodes can vary in their length and consistency, and might be divided into sub-episodes or occur simultaneously so teams can multitask effectively. The framework includes the three higher-order dimensions of action processes, transition processes, and interpersonal processes, which are further divided into 10 lower-order dimensions—mission analysis formulation and planning, goal specification, and strategy formulation under transition processes; monitoring progress toward goals, systems monitoring, team monitoring and backup behavior, and coordination under action processes; and conflict management, motivation and confidence building, and affect management under interpersonal processes. See Figure 6.1 for the full taxonomy and definitions for each of the lower-order processes.

We first used WhisperX [19] to generate text transcripts with word-level timestamps using one audio recording from each group, making manual corrections to the transcription and speaker attribution as necessary. We then trimmed each transcript to the duration of the task (e.g., from the time when a researcher told participants to start through the time they were told their time had elapsed), and categorized each word of the trimmed transcript using the dictionary developed by Mathieu et al. [221]. Following Mathieu et al.'s approach, each occurrence of a word in the CATA dictionary could be assigned to either one of the 10 lower-order dimensions of Marks et al.'s framework; one of the three higher order dimensions (in the case of words relating to multiple lower-order dimensions—e.g., the word *prepare* could relate to mission analysis formulation and planning, goal specification, and strategy formulation); or teamwork overall (if a word mapped to multiple higher-order dimensions). Finally, we counted the frequencies of words belonging to each category in successive 30-second windows with 15 seconds of overlap, chosen to match the approximate time scale of meaningful interaction episodes between group members. The final output of this analysis was a collection of time series for each group member depicting the levels of the higher- and lower-order team processes over the course of the task, as well as for each group as a whole when the respective time series of its members were summed. At the time of writing, transcripts for one group have not yet been analyzed due to audio processing difficulties; levels of team processes have been calculated for the remaining 11 groups.

While analyzing levels of team processes from speech gives some insight into the nature of teams' interaction dynamics during collaboration, nonverbal interactions—such as a team member submitting a prompt to ChatGPT, or participants jointly manipulating a graphical element on the Miro digital whiteboard—cannot be captured using this approach. To gain a better understanding of nonverbal behaviors that occurred during the task, we calculated the frame-by-frame structural similarity index measure (SSIM) [356] for videos of each participant's screen, as implemented in the `scikit-image` library for Python[5]. The SSIM is designed to measure the degree to which two images are perceived as similar to the human visual system, taking into account the effects of noise or distortion which could be ignored when using other measures such as the mean squared error (MSE).

For two images $\mathbf{x}$ and $\mathbf{y}$, the SSIM is given by

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha} \cdot [c(\mathbf{x}, \mathbf{y})]^{\beta} \cdot [s(\mathbf{x}, \mathbf{y})]^{\gamma} \tag{6.1}$$

where $l(\mathbf{x}, \mathbf{y})$ compares the *luminance* of the images, $c(\mathbf{x}, \mathbf{y})$ compares their *contrast*, and $s(\mathbf{x}, \mathbf{y})$ compares their *structure*, with α, β, and γ weighting their relative importance. The

[5] `https://scikit-image.org/docs/0.24.x/api/skimage.metrics.html#skimage.metrics.structural_similarity`

TABLE 1
Taxonomy of Team Processes

Process Dimensions	Definition	Previous Research on Team Processes
Transition processes		
Mission analysis formulation and planning	Interpretation and evaluation of the team's mission, including identification of its main tasks as well as the operative environmental conditions and team resources available for mission execution	Fleishman & Zaccaro (1992); Prince & Salas (1993)
Goal specification	Identification and prioritization of goals and subgoals for mission accomplishment	Dickinson & McIntyre (1997); Levine & Moreland (1990); O'Leary-Kelly, Martocchio, & Frink (1994); Prussia & Kinicki (1996); Saavedra, Early, & van dyne (1993)
Strategy formulation	Development of alternative courses of action for mission accomplishment	Cannon-Bowers, Tannenbaum, Salas, & Volpe (1995); Gladstein (1984); Hackman (1983); Hackman & Oldham (1980); Prince & Salas (1993); Stout, Cannon-Bowers, Salas, & Milanovich (1999); Weldon, Jehn, & Pradhan (1991)
Action processes		
Monitoring progress toward goals	Tracking task and progress toward mission accomplishment, interpreting system information in terms of what needs to be accomplished for goal attainment, and transmitting progress to team members	Cannon-Bowers, Tannenbaum, Salas, & Volpe (1995); Jentsch, Barnett, Bowers, & Salas (1999)
Systems monitoring	Tracking team resources and environmental conditions as they relate to mission accomplishment, which involves (1) internal systems monitoring (tracking team resources such as personnel, equipment, and other information that is generated or contained within the team), and (2) environmental monitoring (tracking the environmental conditions relevant to the team)	Fleishman & Zaccaro (1992)
Team monitoring and backup behavior	Assisting team members to perform their tasks. Assistance may occur by (1) providing a teammate verbal feedback or coaching, (2) helping a teammate behaviorally in carrying out actions, or (3) assuming and completing a task for a teammate	Dickinson & McIntyre (1997)
Coordination	Orchestrating the sequence and timing of interdependent actions	Brannick, Prince, Prince, & Salas (1992); Brannick, Roach, & Salas (1993); Fleishman & Zaccaro (1992); Zalesny, Salas, & Prince (1995)
Interpersonal processes		
Conflict management	Preemptive conflict management involves establishing conditions to prevent, control, or guide team conflict before it occurs. Reactive conflict management involves working through task and interpersonal disagreements among team members	Cannon-Bowers, Tannenbaum, Salas, & Volpe (1995); Gladstein (1984); Jehn (1995); Pace (1990); Simons, Pellad, & Smith (1999); Simons & Peterson (2000); Smolek, Hoffman, & Moran (1999); Tjosvold (1985); Van de Vliert, Euwema, & Huismans (1995)
Motivation and confidence building	Generating and preserving a sense of collective confidence, motivation, and task-based cohesion with regard to mission accomplishment	Fleishman & Zaccaro (1992)
Affect management	Regulating member emotions during mission accomplishment, including (but not limited to) social cohesion, frustration, and excitement	Cannon-Bowers, Tannenbaum, Salas, & Volpe (1995)

Figure 6.1: Marks et al.'s [218] taxonomy of team processes.

SSIM is strictly positive and at most 1 when the images being compared are identical. In the case of our video data, values for the SSIM are highest when a participant is not actively interacting with any tools on their computer, and are lowest when a participant switched tools (e.g., changing their foreground window from Miro to ChatGPT or the Zoom meeting). To reduce the time required for computing the SSIM, videos were re-encoded from 1080p 30fps to 360p 5fps prior to calculation. We then calculated the SSIM on a frame-to-frame basis using the re-encoded video, and averaged these values for each second of the recording.

Due to problems with video recording, the SSIM calculation was omitted for two participants.

6.3.4 Individual EEG Metrics of Attention

Power spectral density (PSD), or the ratio of signal power to signal frequency, is the most frequently used EEG measure to assess mental workload and task engagement [152]. The PSD of the alpha frequency band from the occipital and parietal lobe and the PSD of theta bands from the frontal lobe are the indicators most frequently used in literature. A reduction in the PSD of the parietal alpha bands and an increase in the PSD of the frontal theta bands have been observed when task difficulty or mental workload increases, whereas a decline in beta power seems to indicate the end of a cognitive task.

The Task Load Index (TLI) and Task Engagement Index (TEI) are indices derived from these observations. The Task Load Index is defined as the ratio of the mean frontal midline theta power to the mean parietal alpha power (Equation 6.2), and has been shown to increase with cognitive load during a task [105, 132, 162].

$$\mathrm{TLI} = \frac{\theta}{\alpha} \tag{6.2}$$

The Task Engagement Index is defined as the ratio of mean beta power to (mean alpha power + mean theta power) (Equation 6.3). Task engagement is a positive, excited state that is influenced by cognitive workload. TEI has been shown to increase with task engagement / attention and cognitive load [25, 98, 228, 251].

$$\mathrm{TEI} = \frac{\beta}{\alpha + \theta} \tag{6.3}$$

In the analysis of each participant's individual EEG data, TLI and TEI are used to assess different phases of task engagement and cognitive load during the task, be it during discussion with other team members or during individual thinking. This gives us the ability to assess a participant's engagement with the task as well as with the other participants and how these states change over the course of the experiment session.

The PSD band power values for a run are calculated by computing the mean band power for each of the relevant frequency bands (theta, alpha, and beta) individually over 1-second windows. As both indices only rely on the PSD of the frontal and parietal lobe, we can reduce the number of relevant channels down to 5 (Fz, Oz, PO7, Pz, PO8). This leaves us with 1250 samples per 1-second window (250 samples for 5 channels) and approximately 3600 windows per participant per 1-hour experiment session. The PSD is calculated using MNE's [111] *compute_psd* function. To reduce electric noise and outside interference, we only use the relevant frequencies for each band and ignore all frequencies over 30 Hz. Using these band power values, we then calculated the TLI and TEI for each window and individual participant.

PSD, TLI, and TEI calculations were omitted for two participant in different groups of four whose EEG recordings failed to save. The TLI calculation was also omitted for another participant due to poor signal quality in the Fz electrode.

6.3.5 Team Brain Metrics

To measure the collective brain dynamics of our group members while they were collaborating, we employed four analysis methods which have been used in prior work to relate group dynamics in EEG or other physiological data with levels of team processes: wavelet transform coherence (WTC) [64, 122]; the driver-empath model of team dynamics [117]; mutual information [315]; and multidimensional recurrence quantification analysis (MdRQA) [85].

EEG Preprocessing

Before conducting the group analysis of our EEG data, we first performed preprocessing to remove noise and restrict our analysis to frequencies of interest. The start and end times of the task were determined for each participant using the WhisperX transcripts of their respective videos; these were then used to determine the corresponding indices of each participant's EEG time series by 1) finding the UNIX timestamp of the frame of the video when the task started, 2) finding the corresponding LabRecorder timestamp (which occurs at the same index in the OBS LSL stream), and 3) finding the sample of the EEG LSL stream with the corresponding timestamp. Each EEG time series was trimmed to 9000 samples before and after the task to mitigate any edge effects, then filtered from 0.1–40 Hz (as recommended in [359]). Filtering was performed using EEGLAB's default filter, which is a single-pass zero-phase finite impulse response filter using a windowed sinc function kernel with a wide transition band to limit artifacts and a variable filter order for low- versus high-pass to attenuate the stopband optimally, and compares favorably to default filters in other packages [69]. Finally, the filtered EEG data was trimmed to the duration of the task. Note that for some groups, one of the recording computers wrote data more slowly at a lower effective sampling rate as a result of high usage of system resources, resulting in less data recorded than the full duration of the task.

For all analyses below, data for all participants was truncated to length of the shortest participant stream, and analyses were performed on successive 30-second windows with 15 seconds of overlap. For two groups, the EEG streams from two participants were lost, and were omitted from analysis. Furthermore, two groups had streams from two participants which terminated prematurely 20 minutes after the start of the task; applicable analyses were performed using data from all participants truncated to this shorter length, as well as the full duration of the task with streams from these participants omitted. Finally, four groups had one participant whose data had one to two excessively noisy channels, which were omitted from the analysis on an individual basis (that is, data from these channels was included in the group analysis from participants for whom they were clean).

Wavelet Transform Coherence

Wavelet transform coherence (WTC) is an analysis technique measuring the local correlation in signal power between two or more signals across various frequency bands and time intervals. It is commonly used in hyperscanning studies by the social and cognitive neuroscience communities for measuring inter-brain synchrony [122, 64], which has been commonly observed during group collaboration across a wide variety of tasks and contexts [64, 210, 201, 245, 249, 284] in multiple neurophysiological recording modalities, including fMRI, EEG, and fNIRS. In contrast with other methods of calculating synchrony such as cross-correlation and Fourier transform coherence, WTC is especially suitable for assessing synchrony in physiological data because the its time resolution varies with frequency—with higher resolution with increasing frequency—rather than being fixed for all frequencies, and can thus better account for the non-stationarity of most biological signals and more appropriately model time-varying naturalistic interpersonal interactions.

The ability for WTC to optimize its resolution of different frequency components is due to the use of wavelets in the underlying continuous wavelet transform (CWT). A *wavelet* is a function with zero mean that is localized in both time and frequency [113], and acts as a bandpass filter when convolved with the signal of interest. By convolving the signal with a series of wavelets that are stretched in time to different scales, the power of the signal can be optimally localized across time and frequency. One family of wavelets, the Morlet wavelet (a sine function windowed with a Gaussian), is typically used in wavelet analysis, and is defined as

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\frac{1}{2}\eta^2} \tag{6.4}$$

where ω_0 is dimensionless frequency and η is dimensionless time. Empirically, choosing $\omega_0 = 6$ has been shown to provide a good balance between time and frequency localization, and is the default value used in Grinsted's [113] `wavelet-coherence`[6] and Hu & Si's [142]

[6] `https://grinsted.github.io/wavelet-coherence/`

`Multiple Wavelet Coherence`[7] MATLAB toolboxes used in our analysis.

At different scales $s = \frac{\eta}{t}$, the CWT of a time series x_n ($n = 1, ..., N$) with uniform time steps δt is given by

$$W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n'=1}^{N} x_{n'} \psi_0 \left[(n' - n) \frac{\delta t}{s} \right] \tag{6.5}$$

where the wavelet ψ_0 has been normalized to have unit energy. The wavelet power of the signal is given by the squared magnitude of the CWT $|W_n^X(s)|^2$. Because the transform can introduce edge effects, a cone of influence (COI) where such effects can occur is typically omitted from further analysis.

Extending to comparisons between two signals, the cross-wavelet transform XWT of two time series x_n and y_n is defined as $W^{XY} = W^X W^{Y*}$, where $*$ denotes the complex conjugate. The cross-wavelet power, $|W^{XY}|$, indicates when the signals exhibit high common power in the same frequency range simultaneously.

Finally, the wavelet transform coherence, which provides a measure of the localized cross-correlation between the two signals, is given by

$$R_n^2(s) = \frac{|S\left(s^{-1} W_n^{XY}(s)\right)|^2}{S\left(s^{-1}|W_n^X(s)|^2\right) \cdot S\left(s^{-1}|W_n^Y(s)|^2\right)}, \tag{6.6}$$

where S is a smoothing operator operating along both the time and scale dimensions. Like the typical normalized correlation coefficient, $R_n^2(s)$ ranges from 0 to 1, with a value of 0 indicating no coherence and a value of 1 indicating identity. See [6] for further discussion of the more general multivariate case with more than two time series.

We used Grinsted's [113] `wavelet-coherence` MATLAB toolbox for our analysis (as well as Hu & Si's [142] `Multiple Wavelet Coherence` toolbox for some preliminary analyses integrating data from all group members simultaneously), using default parameters: $Dj = 1/12$ for 12 sub-octaves per scale octave; $S0$ the minimum wavelet scale was double the signal period; and Morlet wavelets were used as the "mother" wavelet of the wavelet family used for the transform. For each 30-second window of (centered) data, we first calculated the mean coherence for the δ (0.5–4 Hz), θ (4–7 Hz), α (7–13 Hz) and β (13–30 Hz) frequency bands for each channel across each pair of participants within the group, omitting all values in the cone of influence. We then calculated the coherence for each frequency band both for all eight channels and the subset of five channels (Fz, Oz, PO7, Pz, PO8) used for the individual analyses above by averaging the subject-pair coherence values for the relevant channel sets, resulting in two collections of four time series (one per frequency band) for each group of participants.

[7] `https://figshare.com/articles/code/Matlab_code_for_multiple_wavelet_coherence_and_partial_wavelet_coherency/13031123`

The Driver-Empath Model of Teamwork

Guastello & Peressini [117] developed the driver-empath model of teamwork to better capture the nonlinear interdependent nature of team dynamics as they arise in group behavior or physiological activity. The authors build on prior work modeling teams as nonlinear dynamical systems, starting with the assumption that the time series originate from low-dimensional chaotic systems—which have been shown to exhibit self-organizing behavior when coupled [20, 224]—and provide several improvements over existing approaches such as phaseclustering or multilevel correlation analysis. Unlike these other approaches for quantifying synchrony, the driver-empath model is able to separate autocorrelational and cross-correlational effects (and thus determine whether cross-correlations between group members are real and not spurious), account for asymmetry in influence among group members, and ultimately provide both individual and group-level metrics for describing the synchrony of the team.

The first step in computing the metrics for the driver-empath model is calculating the prototype matrix $\mathbf{P}$, which contains information about the group's effects on individuals. The matrix is populated by fitting linear or nonlinear Granger causality regression models describing the directional influence of each group members' time series on the others', with the autocorrelations r_{ii} of physiological time series on the diagonal and transfer coefficients r_{ij} on the off-diagonal entries.

Table 6.3: Example prototype matrix for the driver-empath model from [117], calculated for a team with four members. (AR represents autoregression.)

	To				
From	P_1	P_2	P_3	P_4	Driver Score
P_1	AR_1	R_{12}	R_{13}	R_{14}	$\sum R_{1i}^2$
P_2	R_{21}	AR_2	R_{23}	R_{24}	$\sum R_{2i}^2$
P_3	R_{31}	R_{32}	AR_3	R_{34}	$\sum R_{3i}^2$
P_4	R_{41}	R_{42}	R_{43}	AR_4	$\sum R_{4i}^2$
Empath Score	$\sum R_{i1}^2$	$\sum R_{i2}^2$	$\sum R_{i3}^2$	$\sum R_{i4}^2$	

The sums of squared coefficients of the rows and columns are then calculated to produce driver scores and empath scores respectively. The person with the largest row total is the *driver* of the group, the team member who exerted the strongest influence on the others. The person with the largest column total is the *empath*, the team member who was most receptive to the influences of others. The driver has minimal impact on the synchrony between group members unless others in the group are responding, so connections with the empath reflect the strongest synchronization of any one person with the others. This group-level effect is captured by the *synchronization coefficient* S_E (synchronization with the empath).

To calculate the synchronization coefficient S_E, first the empath is identified as above, then the column of coefficients for the empath is removed from $\mathbf{P}$. The empath's autocorrelation is dropped, and the remaining coefficients become a column vector $\mathbf{V}'$. Then the empath's row is removed from $\mathbf{P}$, leaving a square matrix $\mathbf{M}$. Then a weight vector $\mathbf{Q}$ is calculated with

$$\mathbf{Q} = \mathbf{M}^{-1}\mathbf{V}'. \tag{6.7}$$

Finally, the synchronization coefficient S_E is given by

$$S_E = \mathbf{V}'\mathbf{Q}. \tag{6.8}$$

S_E is approximately normally distributed, and does not have bounded lower or upper limits.

We adapted the `ProcPlayer`[8] R script from Guastello & Peressini to calculate the driver and empath scores of each participant as well as the synchronization coefficient for each group where all data was available. (Due to the causal nature of this analysis, two groups where a participant's EEG data failed to record—and thus missing information about the directional influences between group members—were omitted.) This processing script from the authors includes four models with different theoretical origins shown to accurately capture autonomic synchrony in groups [116] which can be used to calculate the $\mathbf{P}$ matrix. Each models the time series of a target participant as a function of the influence of past values from the time series of both the target participant and another group member at a particular lag j.

The first such model is a linear model, which has been shown to sufficiently characterize pairwise levels of synchrony over relatively short time intervals in homogeneous conditions [117]. The model is given by

$$\hat{X}_n = \beta_0 + \beta_1 X_{n-j} + \beta_2 P_{n-j}, \tag{6.9}$$

where $\hat{X}_n$ is the data of the participant of interest at time point n (to be predicted); X_{n-j} is data from the participant of interest, lagged j time points; and P_{n-j} is data from another participant influencing the participant of interest, likewise lagged.

The second is a nonlinear model (Nonlinear Model 1; NL1), which uses a double exponential structure of the form

$$\hat{X}_n = \alpha e^{\beta X_{n-j}} + e^{\delta P_{n-j}}. \tag{6.10}$$

This model was shown experimentally to capture transfer effects between group members more often, and were better correlated with psychological variables and other important effects when applied to physiological data [115].

The third is a nonlinear logistic map model (Nonlinear Model 2; NL2), used by [333] to explore the coordination of behavior with respect to team members' internal psychological states. In this model, each member of a dyad acts as a control parameter modulating the dynamics of their partner, leading to the emergence, maintenance, and disruption of

[8] `https://academic.mu.edu/peressini/synccalc/slinks.htm`

behavioral synchrony and similarity in internal states including autonomic arousal, mood, and motivation. This model represents the simplest structure able to produce these dynamic features, and is given by

$$\hat{X}_n = P_{n-j} X_{n-j}(\alpha - \beta X_{n-j}). \tag{6.11}$$

Finally, the fourth model (Nonlinear Model 3; NL3) is an exponential variant of the logistic map of NL2:

$$\hat{X}_n = \alpha P_{n-j} X_{n-j} e^{\beta X_{n-j}} + \gamma. \tag{6.12}$$

For each 30-second window of (centered) data, the we calculated the driver and empath scores and synchronization coefficient on a per-channel basis, using each model included in the `ProcPlayer` preprocessing script described above, using a lag j of one second (typically enough to capture events of interest in biological data [117, 167]). We then calculated the mean of these metrics across all eight channels and the subset of five channels (Fz, Oz, PO7, Pz, PO8) used for the individual analyses by averaging across the relevant channel sets, yielding time series for the driver and empath scores for each participant and a time series of the synchronization coefficient for each group for each model and channel set.

Mutual Information

Mutual information (MI) is a measure of the amount of information shared between two or more variables, and indicates the degree to which they are mutually dependent. A collection of variables with no mutual information are completely independent; that is, their joint entropy is exactly equal to the sum of individual entropies, and no information is shared. Values of mutual information greater than zero indicate that information about one variable can be gleaned through observations of another. In the context of team dynamics, increased mutual information between the neural signals of team members has been associated with shared attention, social coordination, and synergistic interactions [315].

Suppose we have a set of N measurements x (e.g., a participant's EEG recording) in one or more dimensions, which are realizations of a random variable X (e.g., the underlying mental process) with probability density $\mu(x)$. The Shannon entropy of X is given by

$$H(X) = -\int \mu(x) \log \mu(x)\, dx \tag{6.13}$$

where the base of the logarithm determines the unit ("bits" for base-2, "nats" for the natural logarithm, and "bans" for base-10; the natural logarithm will be used for all subsequent calculations).

Suppose we have another set of N measurements y of random variable Y with probability density $\mu(y)$, such that $Z = (X, Y)$ is the space spanned by X and Y, with associated measurements $z = (x, y)$ and joint probability density $\mu(x, y)$. Then the *joint entropy* $H(X, Y)$

of X and Y is given by

$$H(X,Y) = -\iint \mu(x,y) \log \mu(x,y) \, dx \, dy, \tag{6.14}$$

with $H(X,Y) \leq H(X) + H(Y)$. The *mutual information* $I(X,Y)$ between X and Y is then

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \tag{6.15}$$

or, after simplifying,

$$I(X,Y) = \iint \mu(x,y) \log \frac{\mu(x,y)}{\mu_x(x)\mu_y(y)} \, dx \, dy, \tag{6.16}$$

where $\mu_x(x) = \int \mu(x,y) \, dx$ and $\mu_y(y) = \int \mu(x,y) \, dy$ are the marginal densities of X and Y [183].

It is not possible to exactly calculate the entropy and mutual information of real-world data if the underlying probability densities of the measured variables are unknown. Instead, these quantities are often *estimated*, typically by partitioning the measured data into bins to generate a frequency distribution of its values, which is then used to approximate the probability density functions of X and Y. Despite the existence of estimators using adaptive bin sizes to improve accuracy, such methods nonetheless exhibit systematic errors from partitioning the data to compute the estimate, and from approximating probabilities with frequency ratios. Instead, we used the second estimator provided by Kraskov et al. [183], which estimates mutual information using k-nearest neighbor statistics. This estimator is efficient, adaptive to data of different sizes and dimensionality, and minimally biased. It also is invariant to scaling operations on the data, and does not assume Gaussianity, amplitude comparability, or a particular underlying distribution.

Considering $Z = (X,Y)$ the space spanned by X and Y, the max norm is used for calculating the distance between any two observations z; that is, $||z - z'|| = \max\{||x - x'||, ||y - y'||\}$. $\epsilon(i)/2$ is defined as the distance from an observation z_i to its kth neighbor. $\epsilon_x(i)/2$ and $\epsilon_y(i)/2$ are likewise the distances between the same observation and its kth neighbor in the X and Y subspaces, respectively. This information is used to estimate the joint and marginal probability densities of the variables being measured, i.e., the probabilities such that for any observation x_i, y_i, or z_i there is one point at a distance such that there remain $k-1$ points at smaller distances and $N-k-1$ points at larger distances.

$n_x(i)$ and $n_y(i)$ are defined as the number of points with $||x_i - x_j|| \leq \epsilon_x(i)/2$ and $||y_i - y_j|| \leq \epsilon_y(i)/2$, respectively. Then the final estimator for the mutual information between X and Y is given by

$$I(X,Y) = \psi(k) - 1/k - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N), \tag{6.17}$$

where

$$\langle \cdots \rangle = N^{-1} \sum_{i=1}^{N} \mathbb{E}\left[\cdots(i)\right] \tag{6.18}$$

and $\psi(x)$ is the digamma function; see [183] for the full derivation. Note that while the introduction above estimates the mutual information between two random variables, the estimator can be generalized to estimates for more than two variables as well.

We used the implementation of Kraskov et al.'s [183] second estimator provided in the `rmi` R package for our calculation of mutual information, choosing a value of 5 for k to mitigate the tradeoff between the bias and variance of the estimate [103, 304]. For each 30-second window of data, we calculated the mutual information for each pair of participants both for all eight channels and the subset of five channels (Fz, Oz, PO7, Pz, PO8) used for the individual analyses above. For each subject, all relevant channels were considered simultaneously—that is, each observation for each participant was an n-element vector, where n was the number of channels included in the set. We then averaged the subject-pair mutual information values for the relevant channel sets, resulting in two time series (one per set of channels) per group.

Multidimensional Recurrence Quantification Analysis (MdRQA)

Recurrence quantification analysis (RQA) is a technique for measuring the structural and temporal characteristics of nonlinear dynamical systems without a priori assumptions and only a few free parameters. The main goal of the analysis is to quantify *recurrences*, or repeating patterns or states, occurring in time series data. Though the concept of recurrence has a long history in mathematics, the introduction of RQA by Zbilut & Webber [383] in 1992—and subsequently MdRQA [352], its multivariate extension—led to widespread use in several fields and applications, including Earth science [358], economics [257], civil engineering [106, 370], psychology [286], and ergonomics [85, 110, 156, 344]. Work in psychology and ergonomics in particular has examined the relationship between behavioral and physiological synchrony and the quality of group interactions, relating recurrence-derived metrics (most commonly the recurrence rate, or percentage of recurrent observations) to team members' emotional valence, shared visual attention, frequency and understanding of communication, and levels of coordination and other team processes relevant to collaboration. As Eloy et al. [85] reported a negative correlation between the group recurrence rate of team members' brain signals and levels of team coordination, strategy formulation, and affect management from Marks et al.'s [218] team processes framework, we focus on this metric for our analysis as well.

We begin here with a discussion of the univariate case (i.e., ordinary RQA), then extend this to the more general case for multiple variables (MdRQA) we used in our analysis. A key feature of RQA is its reconstruction of a higher-dimensional representation of the system being analyzed through *time-delayed embedding* to uncover higher-order dynamics.

Consider the time series of interest $\mathbf{x}$:

$$\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n), \tag{6.19}$$

where $\mathbf{x}$ is a vector with values x_1 through x_n, representing observations of $\mathbf{x}$ sampled at regular times $t_1, t_1+\Delta t, t_1+2\Delta t, \ldots, t_1+(n-1)\Delta t$. If we know (or have estimated) the true dimension D of the system from which $\mathbf{x}$ is sampled, then we can construct D-dimensional embedded vectors $\mathbf{V}_i$, $(i = 1, 2, 3, \ldots, n)$ of the form

$$\mathbf{V}_i = (x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(D-1)\tau}), \tag{6.20}$$

where each element of $\mathbf{V}_i$ is an element from $\mathbf{x}$, beginning with x_i sampled at time t_i, and subsequent elements are lagged by integer multiples of the *time delay* τ and likewise sampled at multiples of $\tau\Delta t$. The full phase space of the system $\mathbf{V}$ can then be described using $n-(D-1)\tau$ such vectors:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_{n-(D-1)\tau} \end{pmatrix} = \begin{pmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(D-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(D-1)\tau} \\ \vdots & \vdots & & \vdots \\ x_{n-(D-1)\tau} & x_{n-(D-2)\tau} & \cdots & x_n \end{pmatrix}. \tag{6.21}$$

The row index of $\mathbf{V}$ is a measure of time, and each column index corresponds with a dimension in phase space. Thus the row vectors $\mathbf{V}_i$ correspond with a particular snapshot of the dynamics of the phase space of the system measured at time point i, and the column vectors are lagged copies of the signal of interest. The phase space matrix is then used to construct a *recurrence plot*, a (typically) graphical representation of the recurrent dynamics of the system from which the recurrence rate and other associated metrics are derived [82]. In essence, the recurrence plot for a given time series describes repetitions of phase-space values $\mathbf{V}_i$.

Let RP be the 2-dimensional $n-(D-1)\tau \times n-(D-1)\tau$ matrix depicting the recurrence plot of the phase. A point RP_{ij} is considered recurrent if the distance $||\mathbf{V}_i(\mathbf{x}) - \mathbf{V}_j(\mathbf{x})||$ is smaller than some threshold ϵ, termed the *radius*. This is expressed by

$$\mathrm{RP}_{ij} = \theta\left(\epsilon - ||\mathbf{V}_i(\mathbf{x}) - \mathbf{V}_j(\mathbf{x})||\right) \tag{6.22}$$

where $\theta(x)$ is the Heaviside step function.

Creating the recurrence plot thus necessitates computing the pairwise distances for all $\mathbf{V}_i$; Euclidean distance is the typical choice, but any metric distance function may be used. These distances are typically rescaled relative to the maximum distance, with ϵ likewise expressed as a percentage of the maximum distance, for the sake of comparability between time series. Once the recurrence plot has been created, several metrics exist to quantify aspects of the system's dynamics (e.g., the recurrence rate, Shannon entropy, average line length, etc.) [358]. The metric used for our analysis, the recurrence rate, is the percentage of recurrent points in the recurrence plot.

The extension of RQA to multiple time series in MdRQA is accomplished by including these additional variables in the phase space representation of the system. Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_N)$ be a matrix where each column $\mathbf{y}_j$ is one of N time series. Then we construct a matrix $\mathbf{W}$ such that

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_n \end{pmatrix} = \begin{pmatrix} y_{1,1} & y_{2,1} & \cdots & y_{N,1} \\ y_{1,2} & y_{2,2} & \cdots & y_{N,2} \\ \vdots & \vdots & & \vdots \\ y_{1,n} & y_{2,n} & \cdots & y_{N,n} \end{pmatrix} \tag{6.23}$$

where each element $y_{j,i}$ is the value of $\mathbf{y}_j$ sampled at time t_i. Then the phase space reconstruction $\mathbf{V}$ is given by

$$\mathbf{V} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_{1+\tau} & \cdots & \mathbf{W}_{1+(D-1)\tau} \\ \mathbf{W}_2 & \mathbf{W}_{2+\tau} & \cdots & \mathbf{W}_{2+(D-1)\tau} \\ \vdots & \vdots & & \vdots \\ \mathbf{W}_{n-(D-1)\tau} & \mathbf{W}_{n-(D-2)\tau} & \cdots & \mathbf{W}_n \end{pmatrix}. \tag{6.24}$$

The recurrence plot RP is calculated as in RQA.

We first estimated the optimal time delay τ, embedding dimension D, and radius ϵ parameters for each group, with data for all team members considered jointly (i.e., as a matrix $\mathbf{Y}_N$ where N was the number of channels $\times$ the number of participants in the group). Parameters were estimated separately for two sets of channels: the set of all eight EEG channels and the subset of five channels (Fz, Oz, PO7, Pz, PO8) used for the individual analyses above.

The time delay and embedding dimension were estimated using the procedure outlined in [351] and the associated MATLAB code[9]. First, the optimal delay was estimated via the `mdDelay` function by calculating the average (auto) mutual information (AMI) between the data $\mathbf{Y}_N$ and a copy of itself lagged by a particular value of τ. This procedure was repeated for $\tau = 1, \ldots,$ `maxLag`, with the value for τ minimizing AMI chosen as the optimal delay. Due to the large size of our data, we used `mdDelay` argument values `maxLag` = 1,000, `numBins` = 100, and `criterion` = `'localMin'`, where `maxLag` is the maximum lag tested, `numBins` is the number of bins used for the AMI calculation, and `criterion` is the optimality criterion.

[9] https://github.com/danm0nster/mdembedding

Second, the optimal embedding dimension was estimated using the `mdFnn` function, which determines the percentage of "false nearest neighbors" resulting from possible values of D. The general idea is that an embedding is sufficient if subsequent embedding in higher dimensions does not appreciably alter the relative distances between points in the phase space; if the distances between a point and its nearest neighbor changes by greater than some tolerance when embedded to a higher dimension, or is greater than some absolute threshold, the neighbor is labeled false. The first value of D such that the percentage of false nearest neighbors is zero, or else plateaus to some value, is the optimum. For our dataset, the optimal value of D was 1 for all groups, indicating that the dimensionality of our data was already high enough to sufficiently capture the higher-order dynamics of the system, and that no time-delayed embedding was necessary.

Finally, we estimated the optimal radius ϵ, a non-trivial process for our dataset. The prevailing guidance in much of the existing literature is to select a radius such that the recurrence rate is "small" and the dynamics of the system are visible in the recurrence plot, and provide some rules of thumb for doing so—[358] recommends the recurrence rate be kept from 0.1 to 2%; [350] recommends it be 1 to 5%, but states that some stochastic time series might warrant higher values; and [59] recommends choosing a radius yielding a recurrence rate from 2 to 5%, but also notes that the choice of radius must be tailored to the particular dataset being analyzed, as each has its own idiosyncrasies. However, as recurrences can be meaningful at various distance scales in the phase space, we adopted a more data-driven approach used by Yang et al.[370].

The idea is to choose the radius that provides the best discrimination between our real data and surrogate data with comparable spectral qualities with respect to the recurrence metric of interest. This assures that the recurrence dynamics observed are caused by the true underlying dynamics of the system, rather than noise or artifacts. This optimization procedure was performed with R using the `wsyn` and `ParBayesianOptimization` packages, as well as a modified version of the `crqa` package able to leverage multiprocessing to more efficiently calculate the recurrence metrics of our large dataset (specifically regarding the calculation of pairwise distances, which requires $O(n^2)$ memory). We first construct a collection of N phase-randomized surrogates, using the amplitude-adjusted Fourier transform method [296] implemented by the `surrog` function in `wsyn` (which does not require the data to be normally distributed). A surrogate that is *phase-randomized* has the same PSD as the original signal with the phase component of its frequency domain representation randomized, destroying its time-varying structure. All channels of the original data and its surrogates were then z-scored, and the maximum pairwise distance of the z-scored original data was calculated; this distance was subsequently used to rescale the distance matrix for all calculations of the recurrence rate. Finally, we implemented an optimization routine to determine the optimal radius, using the Taguchi quality loss function from [370] as our scoring function.

Considering a quality characteristic represented by Y

$$Y = |Q_{\text{Surro}} - Q_{\text{Ori}}|, \tag{6.25}$$

where Q_{Surro} is the recurrence rate from surrogate data and Q_{Ori} is the recurrence rate from original data.

The loss $L(y)$ incurred for a given value $Y = y$ is given by

$$\text{QLF} = KV_T \tag{6.26}$$

$$V_T = \frac{1}{n}\sum_{i=1}^{n}(y_i)^2 = \sigma_Y^2 + \mu_Y^2 \tag{6.27}$$

where σ_Y^2 and μ_Y^2 are the variance and mean of y. The optimal radius is the value which maximizes the loss function QLF, maximally discriminating the original data from the surrogate data. Here K is a constant representing cost; we used a value of 100. For a given value of the radius, a the recurrence rate Q_{Surro} for each surrogate and Q_{Ori} for the original data are calculated. Rather than calculate the recurrence rate using all data points, we used a sampling approach to reduce the computational resources required; for the original data and each surrogate, the mean recurrence rate from 60 randomly selected 30-second windows was used for Q_{Ori} and each Q_{Surro}. This scoring function was then used to find the optimal radius via Bayesian optimization, as implemented the in `ParBayesianOptimization` package.

Finally, once the optimal values for the delay τ, embedding dimension D, and radius ϵ were obtained, we used these values to calculate recurrence rates for successive 30-s windows (with 15-s overlap) for each group. As in the optimization procedure above, first the channels of entire dataset for each group were z-scored, then the maximum pairwise distance was calculated, and subsequently used as a scaling constant for the distance matrix of each window. This procedure was performed using all eight channels and the subset of five channels (Fz, Oz, PO7, Pz, PO8) used for the individual analyses above, resulting in two recurrence rate time series per group (one per channel set).

6.3.6 Analyzing the Relationship Between Individual and Team Metrics

We employed several methods to examine the relationships between the different behavioral and neural metrics described above. As the vast majority of our analyses return time series data, we used several techniques suitable for comparing multiple time series. We used dynamic time warping distance and similarity to compare the individual neural metrics (TLI and TEI) of participants within each group, as well as to determine the relationship between each participant's individual neural metrics with the team's group level neural metrics. We used cross-correlation and sliding window correlation to additionally these metrics to our measures of team processes, as well as linear and decision tree regression to examine the relationships both among continuous measures as well as between these measures and team design scores.

Dynamic Time Warping

Dynamic time warping (DTW) is an an elastic method for comparing two sequences. In contrast to pointwise distance metrics such as Euclidean distance, which require one-to-one correspondence between sequences of equal length, elastic distance metrics allow for many-to-one mappings between sequences of potentially different lengths by warping them along the time axis, aligning each point with its most similar neighbor in a local window [169].

We used functions provided by the `DTAIDistance` Python library to calculate DTW distance for our analysis [229]. `DTAIDistance` first calculates the Euclidean distance between each pair of data points. It then generates a *warping path*, a mapping of alignments showing which point of the first time series is aligned with which point of the second series. This can be used to take a closer look at temporal differences between the two time series, and can be especially useful if the lag or difference between the two series varies over time. In all cases, we provide z-scored input sequences, and restrict the search for each point's optimal match to a window of three adjacent points.

In addition to the DTW distances, `DTAIDistance` can calculate a similarity measure for the analyzed time series, which enhances interpretability. This is accomplished by mapping the DTW distance values on the interval $[0, \infty)$ to similarity scores on the interval $[1, 0)$, i.e., $\mathrm{DTW}_{\mathrm{sim}} = e^{-D}$, where D is the DTW distance. Note that this mapping is non-linear; if comparing to the same target signal, the relative difference between time series with scores of 0.8 and 0.9 is smaller than that of series with scores of 0.2 and 0.3. Thus the similarity score magnifies relative differences between signals closer to the target than those farther away.

We calculated the DTW Distance between each individual index for each individual participant and each of the group measures.

Cross-Correlation

In addition to the DTW Distance, we also calculated the normalized cross-correlation for each of the individual indices and the group measures. Cross-correlation is a measure of similarity between two time series as a function of their relative displacement to one another. It can be considered a "sliding inner product"; for two signals x and y each of length n, the cross-correlation between x and y at lag any lag $m \in [-n, n]$ is the inner product between x lagged by m and the original non-lagged signal y. The result is a list of correlation coefficients for each possible lag, with the cross-correlation at lag 0 equivalent to the Pearson correlation.

Analogous to Pearson correlation coefficients, values for the correlation coefficients from cross-correlation are bound between -1 and 1.

Sliding Window Correlation

Due to the long length of the data collected from our experiment sessions, which each contain several thousand data points per participant even after processing, it is unlikely that any of the derived measures would be perfectly correlated. Thus we also employed sliding window correlation for each of the individual and the group measures.

Windowed correlation calculates the Pearson correlation on smaller windows that are slid over the time series, which increases our likelihood of identifying possible subsequences with high correlation. We used a window size of 5 min with a 2.5 min overlap, unless otherwise indicated. Most of our analyses were already windowed to 30-second windows with a 15-second overlap, so this larger window was chosen to get at least 10 data points per window and increase the chance of getting a statistically relevant result. The 50% overlap was chosen to be in line with the windowing of the other analyses. We report the mean correlation coefficient across all windows unless otherwise specified.

Linear Regression Analysis

We also conducted regression analyses in addition to the correlation analyses above. Regression analysis is appropriate for examining how one or more independent variables affect a dependent variable, making it well-suited for exploring the relationship between our different neural measures and the multiple dimensions of team processes extracted through the text analysis.

For our analysis, we used the ordinary least squares linear regressor provided by `scikit-learn` [263] using default parameters. We used the *coefficient of determination*, or r^2, value to evaluate the goodness of fit of our models. The r^2 value indicates how much the variation in the dependent variables predicted by the model is explainable by the independent variables.

Decision Tree Regression

In addition to linear regression, we used decision tree regression, a method that is more adaptable to multiple different independent variables. As with the linear regression, we used the Decision Tree Regressor provided by `scikit-learn` [263]. We initialized the regressor with the maximum tree depth set to 3, minimum samples per split set to 10 and the minimum samples per leaf node to 5. These parameters were chosen to reduce the chance of overfitting on the data, according to the `scikit-learn` documentation. The settings were the same for all evaluations to keep comparability between the different indices and group measures. As with the linear regression, we based our evaluation on the r^2 results. We also examine the coefficients of the decision trees and the tree structure to take a closer look at the impact of the different input variables.

6.4 Results

In this section we outline findings from our analyses exploring the relationships between neural measures of individual and team dynamics during collaboration, as well as their respective relationships with team performance and behavior.

Due to the richness of our dataset and the large number of derived metrics, analysis is still ongoing, and several planned analyses are incomplete at the time of writing—we discuss these further in Section 6.5. Specifically, analyses examining results from the PMERQ as well as the driver-empath model and WTC metrics for team brain dynamics were omitted from this section, though they are described earlier.

6.4.1 Relationship Between Individual and Group Neural Measures

To examine how individual-level EEG measures relate to group-level synchrony metrics, we compared Task Engagement (TEI) and Task Load (TLI) indices to both mutual information (MI) and MdRQA recurrence rate (RR). To enable direct comparison, we re-windowed the individual indices using 30-second windows with 15-second overlap, matching the group measures.

Dynamic Time Warping (DTW) distance analyses showed that, for TEI, most groups exhibited lower distances to MdRQA than to MI (mean 12.8 and 14.2, respectively) and higher similarity (mean .44 and .49, respectively), indicating a closer alignment between individual engagement and group recurrence patterns. This trend was present in 10 of 12 groups, and was statistically significant in a paired t-test ($t(40) = -4.36$, $p < .001$). For TLI, no consistent pattern emerged.

Cross-correlation analyses further supported a link between TEI and MdRQA. Many participants showed modest positive correlations ($r > 0.2$) within a $\pm$5-lag window, with some reaching above 0.5. A corresponding negative correlation between TEI and MI was observed in several cases, though not universally. For TLI, results were more variable, with no consistent pattern across participants.

Together, these results suggest that individual engagement, as measured by TEI, may be more strongly associated with group recurrence than mutual information—though the observed relationships are modest (average $r \approx .23$) and should be interpreted cautiously.

6.4.2 Influences on Team Performance

The distribution of escape room design scores across all groups is shown in Figure 6.2. Overall, team performance varied considerably, with scores ranging from 49.5 to 93.5 out of a possible maximum of 100 (M = 74.8, SD = 14.1). While there were some groups at the extreme ranges of performance, with two groups each in the highest and lowest bins, most groups performed modestly well, reaching scores of ~ 75 or above.

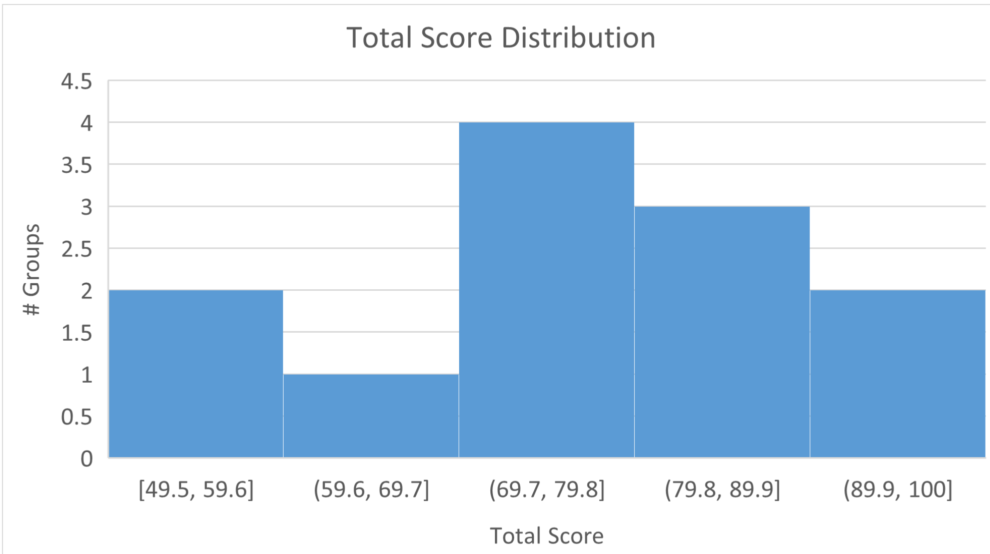


Figure 6.2: Distribution of scores for escape room designs. Scores were grouped into five equal-width bins, ranging from the minimum score (49.5) to the largest possible score (100). The largest score received by any group was 93.5.

Here we discuss the influences of team composition and team processes on the scores for each group's escape room designs.

Impact of Prior Escape Room Familiarity on Team Performance

Although all participants were briefed prior to the task so they would have a baseline familiarity with generative AI and virtual escape rooms, that did not preclude participants from already having some familiarity prior to the experiment. We assessed each participant's familiarity with escape rooms before their experiment session to determine whether it had an impact on team performance; an overview for each group is shown in Figure 6.3.

The design scores for each group and each participant's prior familiarity with escape rooms are provided in Figure 6.4. Familiarity levels could range from (1 - Not at all familiar) to (4 - Very familiar), as in Section 6.2.2, and are listed for each participant (P1 or P2) at each study location (Location 1 - WPI or Location 2 - UniBremen) in the final four columns of the figure.

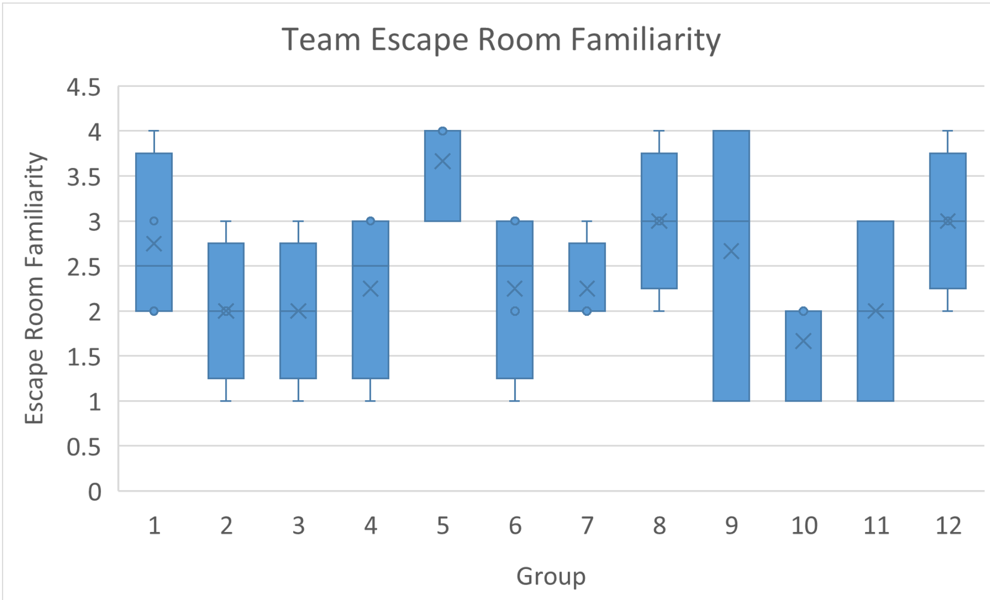


Figure 6.3: Box and whisker plots for each group showing team members' prior familiarity with escape rooms.

We had initially assumed that teams which had one or more members who were "Very familiar" with escape room games would score higher than teams with a lower familiarity, with members who had higher familiarity providing guidance and leadership for those with less. However, this did not bear out in our data: while the a team with the highest familiarity (a value of 4) achieved the highest score and the team with the lowest familiarity (a score of 2) achieved the lowest score, there was no clear consistent relationship between prior familiarity and team performance on the task. This is further illustrated in Figure 6.5, which shows that teams with a maximum familiarity of 3 or 4 achieved a similar range of scores. Indeed, Group 5 was one of the lowest scoring teams, despite having a mean familiarity of 3.7, while Group 3 achieved the second-highest score despite a relatively low mean familiarity of 2.

The absence of a strong relationship is supported by a relatively low Pearson correlation of $r = .25$ between a team's maximum escape room familiarity and the score of their design. Fitting a linear regression model to predict design score as a function of maximum familiarity and mean familiarity yielded values of $r^2 = .09$ and $r^2 = .006$, respectively, further corroborating the minimal effect of prior familiarity with escape room familiarity on team performance.

An additional point of observation is the mediating impact of participant location on familiarity with escape rooms, with participants at L1 having a mean familiarity of 2.8 and participants in L2 having a mean familiarity of 2.1. This is likely indicative of cultural differences in the prevalence of or exposure to escape rooms at each of the study locations.

Group	Total Score	Escape Room Familiarity			
		Location 1		Location 2	
		P1	P2	P1	P2
1	93.5	4	3	2	2
2	61.25	2	3	2	2
3	91.75	2	2	3	1
4	74.5	3	1	3	2
5	50.5	4	4	3	
6	88.75	3	3	2	1
7	71.5	2	3	2	2
8	80.5	4	2	3	3
9	73.75	4	3	1	
10	49.5	2	2	1	
11	85	2	3	1	
12	77.5	3	3	2	4

Figure 6.4: Escape room design scores and prior familiarity with escape rooms for all groups. Design scores could range from 0 to 100, and are color-coded based on the minimum and maximum scores that were achieved—49.5 (the minimum) maps to red, and 93.5 (the maximum) maps to green, with intermediate values interpolated. Familiarity levels are provided for each participant (P1 or P2) at each study location (Location 1 - WPI or Location 2 - UniBremen) in the final four columns: 1 - Not at all familiar; 2 - Somewhat familiar; 3 - Moderately familiar; 4 - Very familiar. Values are likewise color-coded, with 1 mapping to red and 4 mapping to green. Blank cells indicate groups with fewer participants.

Relationship Between Team Performance and Levels of Team Processes

To explore whether team communication patterns were related to design outcomes, we examined the relationship between escape room design scores and the three higher-order team process dimensions from Marks et al.'s taxonomy outlined in Section 6.3.3: transition processes (mission analysis, goal specification, strategy formulation); action processes (goal monitoring, team monitoring, systems monitoring, coordination); and interpersonal processes (conflict management, motivation building, affect management). For each group, counts of utterances in each category were summed across the task duration (see Table 6.4 and Figure 6.6).

When computing Pearson correlations across all 11 groups with transcripts (Table 6.5), we observed a strong negative relationship between interpersonal processes and team performance ($r = -.62$), as well as a moderately negative correlation between transition processes and performance ($r = -.47$). These findings suggest that higher frequencies of interpersonal and planning-related communication may be associated with lower design scores for our task. However, these results appear to be strongly influenced by two statistical outliers: Group 3, which communicated relatively little but achieved a high score, and Group 5, which showed high overall communication but performed poorly.

To assess the robustness of these correlations, we repeated the analysis with Groups 3

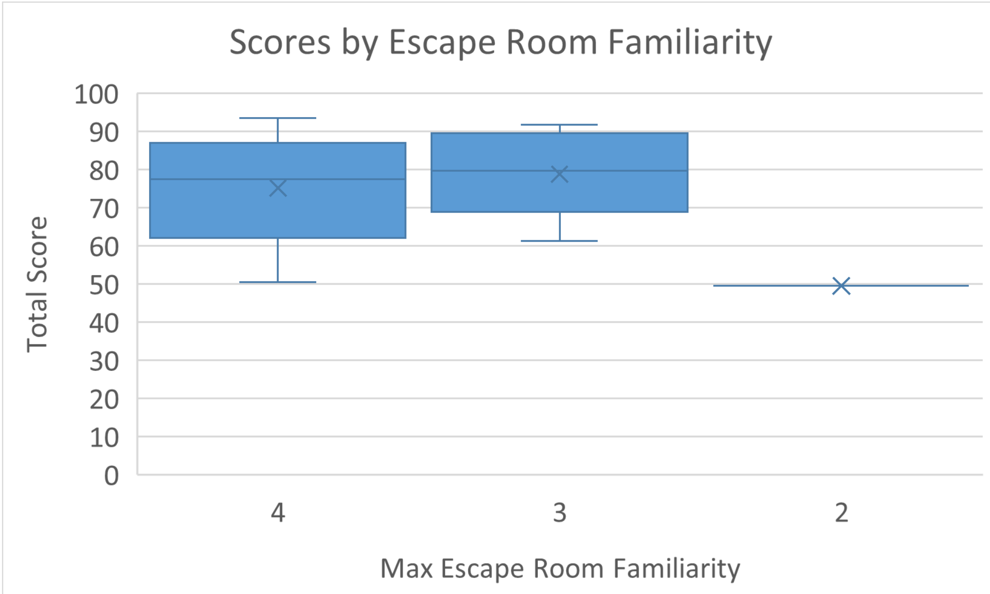


Figure 6.5: Box and whisker plots illustrating the relationship between design scores and the maximum prior familiarity with escape rooms for each team.

and 5 excluded. This adjustment substantially weakened the negative correlations for the interpersonal ($r = -.24$) and transition ($r = -.17$) process dimensions. Interestingly, the correlation between action processes and performance shifted from slightly negative ($r = -.12$) to moderately positive ($r = .53$), suggesting that task-focused communication such as coordination and monitoring may support higher performance, while excessive planning or interpersonal negotiation could potentially hinder progress under time-constrained conditions.

These results do indicate that team processes play a role in governing team performance during collaboration, corroborating prior work. However, due to our small sample size and the potential influence of outliers, these results should be interpreted with caution. Further study with a larger sample is likely necessary for more robust conclusions. Furthermore, while team processes are correlated with performance, the nuances of these relationships are unclear. The negative association between interpersonal processes and performance, for instance, might indicate that teams achieve lower performance *because of* their levels of these processes, or that teams which, e.g., felt time pressure more acutely or had members with incompatible personalities were likely to require more frequent conflict and affect management, and *also* exhibit lower performance. It is unknown whether our findings reflect phenomena specific to our task or collaboration more generally.

The fact that the relationship changes drastically with the removal of two outliers suggests that important aspects of collaboration during our task may be unaccounted for. Since our current analysis is largely restricted to examining team processes as they emerge through verbal communication, future analyses will also examine *nonverbal* interactions and their

Table 6.4: Counts of the utterances for each higher-order dimension of Marks et al.'s taxonomy of team processes [218] and escape room design scores for each group. Utterances were summed across group members for each category; the total number of such utterances for all group members is shown in the "Total" column.

Group	Transition	Action	Interpersonal	Total	Score
1	530	684	986	2688	93.5
2	594	578	1142	2834	61.25
3	178	246	344	878	91.75
4	288	358	614	1546	74.5
5	572	764	1482	3328	50.5
6	430	656	696	2232	88.75
7	544	394	854	2346	71.5
8	362	606	720	2050	80.5
9	330	480	996	2198	73.75
10	426	366	798	1842	49.5
12	306	378	678	1534	85

Table 6.5: Pearson correlation between higher-order team process dimensions and design scores.

	Transition	Action	Interpersonal	Total
All Groups	-.48	-.12	-.62	-.45
No Outliers	-.17	.53	-.24	.05

relationship with performance and team processes.

6.4.3 Relationships Between Individual Neural Measures and Behavior

While we will devote more focus to the relationships between group measures in this chapter, we summarize results exploring the relationship between individual neural measures and behavioral measures here to provide a full picture.

Relationship Between Individual Neural Indices and Team Performance

Building on our earlier analysis of team performance, we next examined whether neural indices of task engagement and cognitive load measured from the brain activity of individual participants were related to the quality of the escape room designs produced by each group. Specifically, we computed mean values for the Task Engagement Index (TEI) and Task Load Index (TLI) across the duration of the task for each participant, and then averaged these within groups to obtain group-level TEI and TLI values.

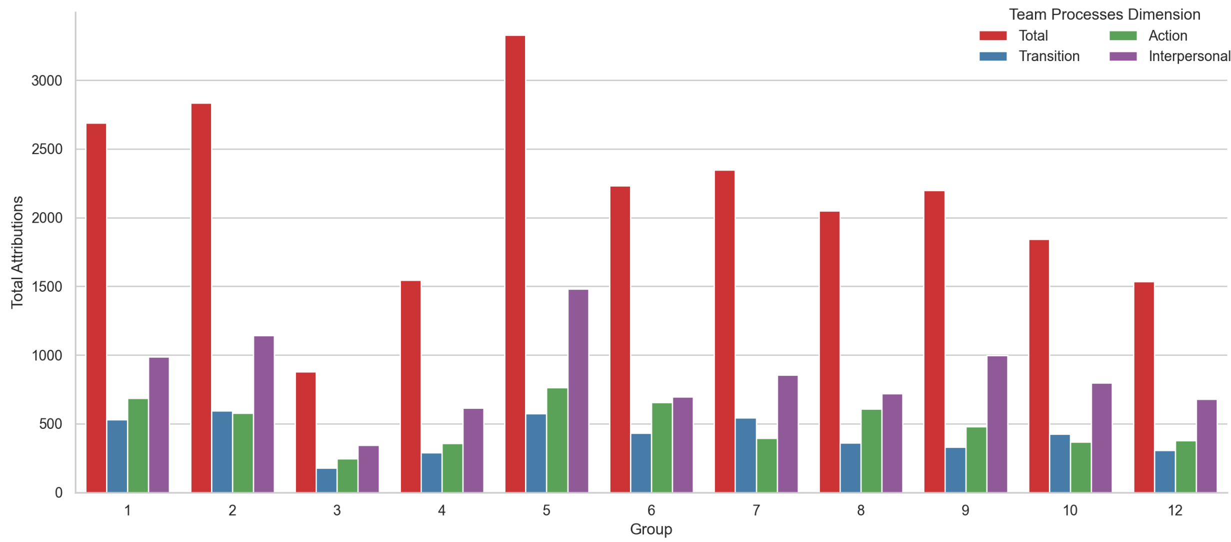


Figure 6.6: Counts of higher-order team process dimensions per group.

We first assessed correlations between each index and the groups' design scores. Both indices showed small but positive associations with design quality, with TLI ($r = .34$) showing a slightly stronger relationship than TEI ($r = .26$). These results suggest a modest link between increased cognitive demand and higher design performance, potentially indicating that more cognitively active teams engaged more deeply with the task.

To further explore predictive relationships, we conducted regression analyses using both linear and decision tree models (Figure 6.7). Across models, TEI and TLI outperformed measures of prior experience (e.g., average or maximum escape room familiarity) in predicting design scores for at least one of the regression techniques. The strongest individual predictor was TEI, with a decision tree r^2 of .13. When both indices were combined, the linear model showed modest improvement ($r^2 = .11$), while the decision tree result remained unchanged.

Incorporating maximum prior experience into the model yielded a further increase in predictive accuracy, particularly in the linear regression, where the combined model (TEI + TLI + maximum experience) reached an r^2 value of .18. By contrast, average prior experience continued to show little predictive value. These findings suggest that both cognitive activation during the task and the presence of at least one experienced group member may play a role in successful design outcomes, although the overall explanatory power remains limited.

Taken together, while no strong predictive relationship emerged, the results point to a small but consistent role for cognitive engagement and task load in shaping collaborative creative performance. These findings complement earlier behavioral analyses by suggesting that more cognitively involved teams may be better positioned to generate high-quality designs under time constraints, though again, more data is necessary for definitive conclusions.

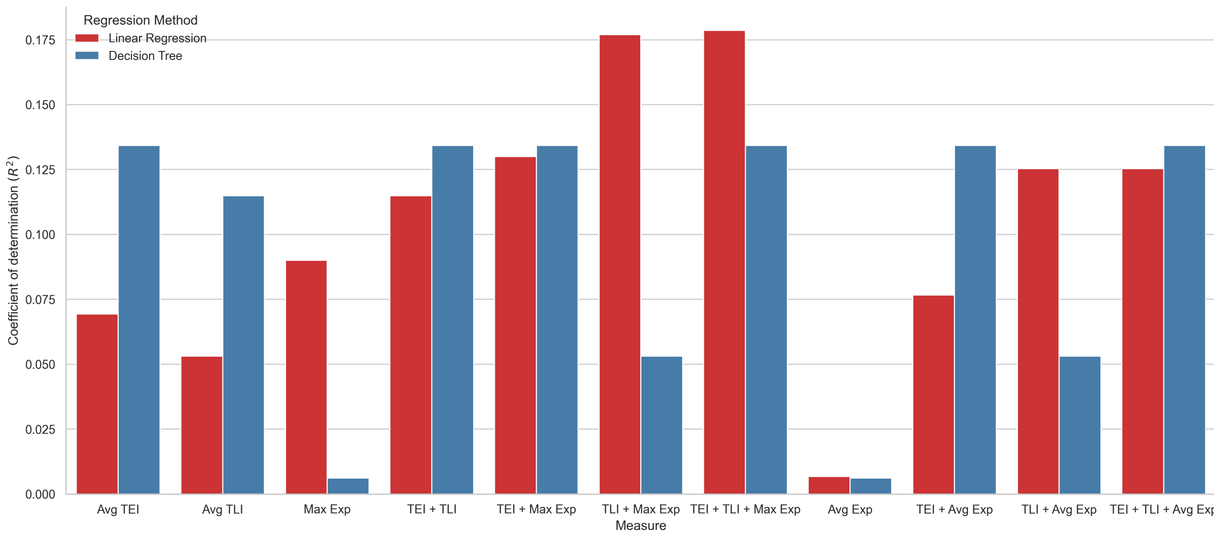


Figure 6.7: r^2 values from regression analyses modeling escape room design scores as a function of individual EEG indices and prior familiarity with escape rooms.

Relationship Between Individual Neural Measures and Levels of Team Processes

While uncovering a link between EEG indices and team performance could be useful for monitoring or evaluating team performance and progress during creative collaboration, uncovering relationships with team processes could prove even more useful, as they provide a more holistic picture of team function and dynamics. To explore this possibility, we followed the same procedure as our prior analysis from Section 6.4.2 investigating the relationship between design scores and levels of team processes.

Neural Indices and Team Processes Using windowed correlation and regression analyses, we compared TEI and TLI to transition, action, and interpersonal team process categories. Correlations across individual participants were generally weak, with most near zero and only a few exceeding ±.2. Regression models offered clearer insight: decision tree regressors consistently outperformed linear models (Figures 6.8 and 6.9), with group-level r^2 values ranging from .15–.25 for TEI and .12–.22 for TLI. Interpersonal processes emerged most frequently as predictors of TEI, while TLI had a more evenly distributed set of predictors (Tables 6.6 and 6.7).

At the team level, we observed modest negative correlations between average TLI and overall team processes, strongest for transition processes ($r = -0.39$; Table 6.8)). Removing outlier teams reduced these effects, suggesting sensitivity to group composition (Table 6.9). Taking the trends at face value would seem to suggest that increased levels of communication and planning among the team lead to reduced workload. However, further analysis of nonverbal coordination and interaction between team members is necessary for a fuller

Table 6.6: TEI decision tree regression coefficients for higher-order team process dimensions.

	Overall Teamwork	**Transition**	**Action**	**Interpersonal**
Times Picked	4	6	8	15
Avg Coef	.16	.18	.27	.40
Max r^2	.42	.45	.30	.47
Mean r^2	.26	.26	.17	.22

picture.

Frequency Bands and Team Processes We also evaluated how EEG frequency bands (theta, alpha, and beta) related to team processes (Figures 6.10, 6.11, and 6.12). Beta band features yielded the highest predictive accuracy across models (r^2=.23 for decision trees), slightly outperforming TEI and TLI. While beta activity was evenly associated across team processes, theta and alpha bands were more closely linked to interpersonal dynamics.

Regression results further indicated that alpha and beta activity were best predicted by action processes (mean r^2 = .29 and mean r^2 = .27, respectively), and theta by interpersonal processes (r^2=.25) (Tables 6.10, 6.11, and 6.12). Out of all three frequency bands, beta activity had the highest mean r^2 value (r^2 = .25) for overall team processes, in line with earlier results for TEI.

Taken together, these analyses had limited predictive power, and do not indicate there was a strong overall relationship between the neural indices and team processes. The most prominent relationship was a moderate negative correlation between TLI and transition processes, with a comparable negative correlation with overall teamwork, which could indicate engaging in early planning and goal setting result in lower team workload, but further investigation is necessary to substantiate this. Regression models also suggest similar moderate relationship between team communication and beta band activity, which are plausible given beta activity is associated with active thinking and concentration. The differences in r^2 values between the regression models of the frequency bands additionally indicate that the power of different frequency bands could be more responsive to different team processes, though the relationship is modest and varies between groups.

6.4.4 Relationships Between Group Neural Measures and Behavior

In this section, we explore the relationships between team behavior and the collective brain dynamics of team members. Due to the size of our dataset and time constraints, we only include the recurrence rate (RR) from MdRQA and mutual information (MI) as our group

Table 6.7: TLI decision tree regression coefficients for higher-order team process dimensions.

	Overall Teamwork	**Transition**	**Action**	**Interpersonal**
Times Picked	7	8	9	9
Avg Coef	.22	.25	.28	.26
Max r^2	.31	.38	.26	.32
Mean r^2	.21	.19	.15	.18

Table 6.8: Pearson correlation between higher-order team process dimensions and mean individual neural measures (all groups).

	Transition	**Action**	**Interpersonal**	**Total**
TEI	-.16	-.25	-.12	-.22
TLI	-.39	-.25	-.19	-.32

Table 6.9: Pearson correlation between higher-order team process dimensions and mean individual neural measures (outliers removed).

	Transition	**Action**	**Interpersonal**	**Total**
TEI	.002	-.12	.13	-.06
TLI	-.13	-.002	.16	-.04

Table 6.10: Theta band decision tree regression coefficients for higher-order team process dimensions.

	Overall Teamwork	**Transition**	**Action**	**Interpersonal**
Times Picked	5	9	7	12
Avg Coef	.20	.25	.24	.31
Max r^2	.28	.29	.30	.51
Mean r^2	.15	.18	.20	.25

Table 6.11: Alpha band decision tree regression coefficients for higher-order team process dimensions.

	Overall Teamwork	**Transition**	**Action**	**Interpersonal**
Times Picked	6	5	8	14
Avg Coef	.18	.21	.25	.36
Max r2	.26	.24	.46	.47
Mean r2	.22	.13	.29	.20

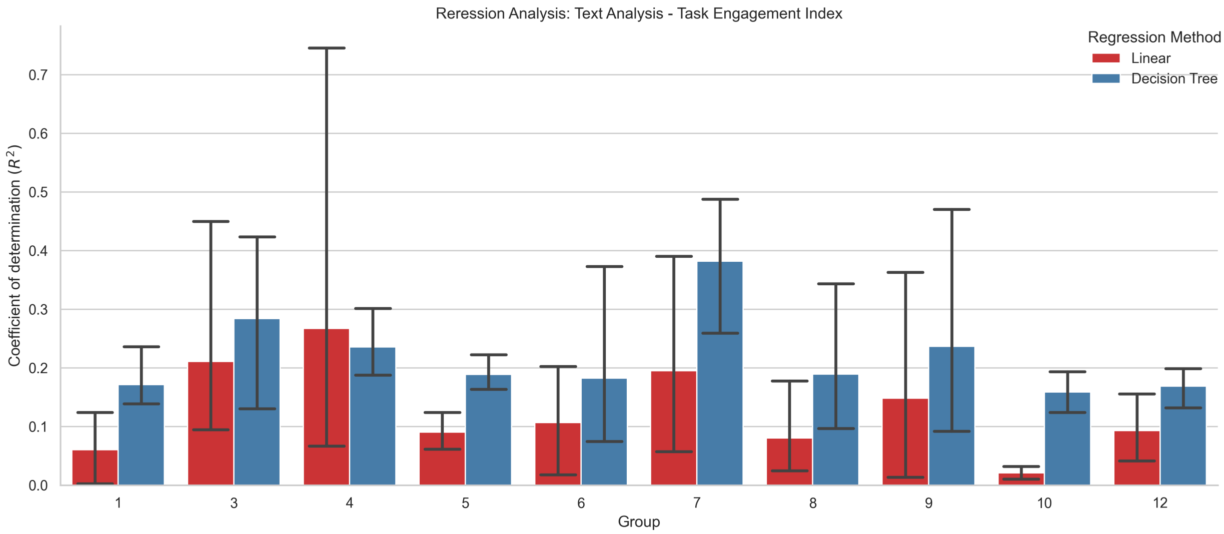


Figure 6.8: Regression modeling TEI as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group.

Table 6.12: Beta band decision tree regression coefficients for higher-order team process dimensions.

	Overall Teamwork	**Transition**	**Action**	**Interpersonal**
Times Picked	8	7	10	8
Avg Coef	.19	.25	.29	.27
Max r^2	.35	.34	.57	.45
Mean r^2	.25	.18	.27	.22

brain measures here; additional analyses incorporating wavelet transform coherence and the driver-empath model are left for future work.

Relationship Between Group Neural Measures and Team Performance

To assess whether group-based neural synchrony measures related to team performance, we analyzed correlations between escape room design scores and the summary statistics (minimum, maximum, mean, and standard deviation) of the MI and RR values for each group values. Of the two, MI exhibited stronger associations with design outcomes. The highest correlation was observed for minimum MI values ($r = .31$), followed by the standard deviation of MI ($r = -.29$), and mean MI ($r = .22$). In contrast, recurrence rate showed a minimal relationship with design quality, with all correlation coefficients below 0.1.

Regression analyses reflected similar patterns. Among all predictors, the standard deviation of MI yielded the strongest predictive performance, with an r^2 of .087 in a linear model

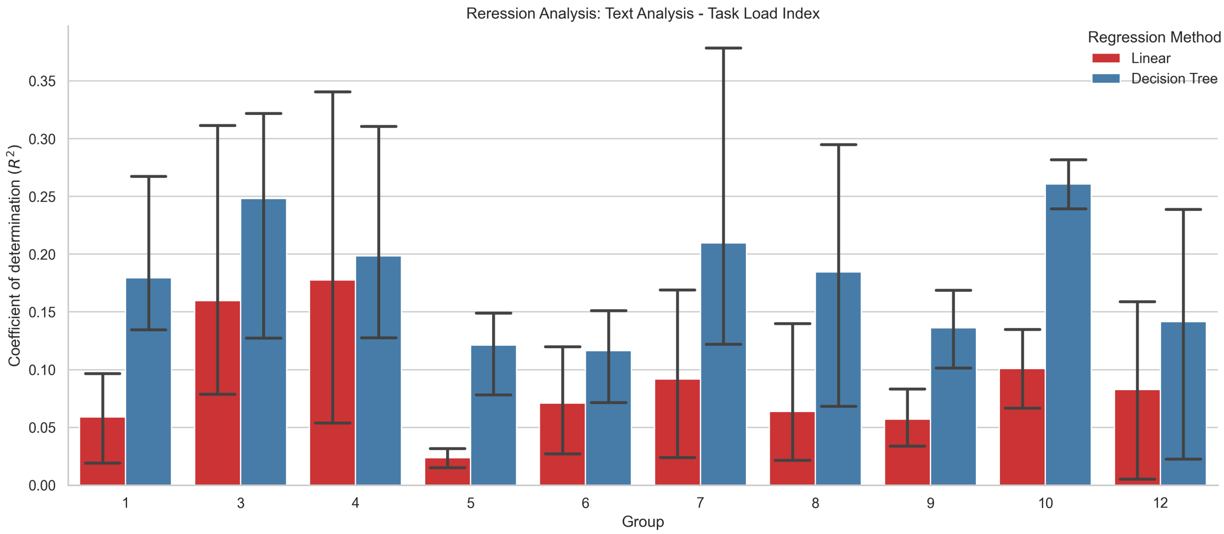


Figure 6.9: Regression modeling TLI as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group.

and .16 using a decision tree. MdRQA values again showed negligible predictive power.

While results suggest a weak relationship between group-level MI and collaborative design quality, these effects are small and should be interpreted cautiously. Given the limited sample size ($n = 12$), further research with larger datasets is needed to determine whether these patterns hold more broadly.

Relationship Between Group Neural Measures and Levels of Team Processes

To explore how group-level neural synchrony relates to team communication dynamics, we applied the same analysis methods as with the individual neural measures. We first performed windowed correlation to measure the similarity of the measures, then created linear and decision tree regression models to predict MI and RR values based on higher-order team process dimensions. Both measures were already computed using the same 30-second sliding window with 15-second overlap as the team process time series, so no additional alignment was required.

Correlation analyses showed minimal direct relationships between team processes and group synchrony. MdRQA RR coefficients ranged from -0.083 to 0.022, while MI correlations were similarly low and consistent across dimensions (-0.058 to -0.051). Regression models, however, yielded modest predictive performance. For RR, the average r^2 was 0.061 (linear) and 0.16 (decision tree), with Group 7 achieving the highest accuracy ($r = .27$) (Figure 6.13). For MI, average r^2 values were .058 (linear) and .15 (decision tree), with Group 3 reaching the top result ($r^2 = .22$) (Figure 6.14).

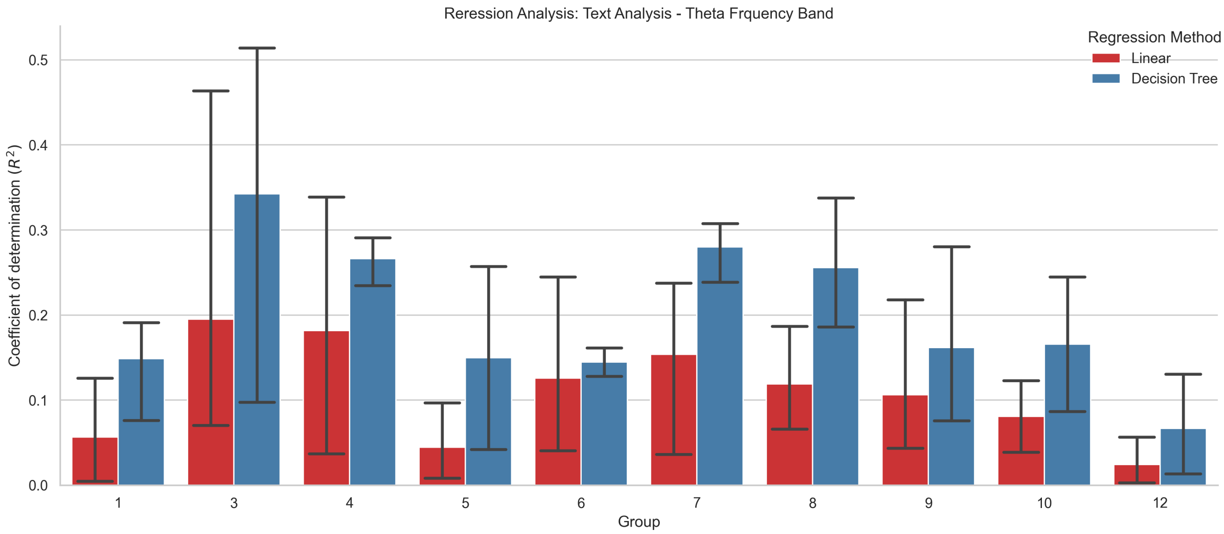


Figure 6.10: Regression modeling theta band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group.

Table 6.13: Pearson correlation between higher-order team process dimensions and mean group neural measures.

	Transition	**Action**	**Interpersonal**	**Total**
MdRQA	-.18	-.045	-.38	-.27
MI	.054	.51	.018	.18

We also examined relationships between overall synchrony and team processes by correlating group mean MI and MdRQA RR values with total utterances in each team process category. Two strong relationships emerged: a negative correlation between mean recurrence rate and interpersonal processes ($r = -.38$), and a positive correlation between mean MI and action processes ($r = .51$) (Table 6.13). After removing outlier groups (3 and 5), these effects strengthened considerably ($r = -.64$ with $p = .06$ and $r = .78$ with $p < .05$, respectively), representing the strongest observed correlations in the study (Table 6.14).

These findings suggest that higher inter-brain recurrence may be linked to reduced interpersonal dialogue form managing conflict, emotions, and motivation, while increased mutual information is associated with task-oriented coordination. While exploratory, these patterns represent highly promising results, corroborating findings from prior work linking RR and MI with collaborative behavior at the group level.

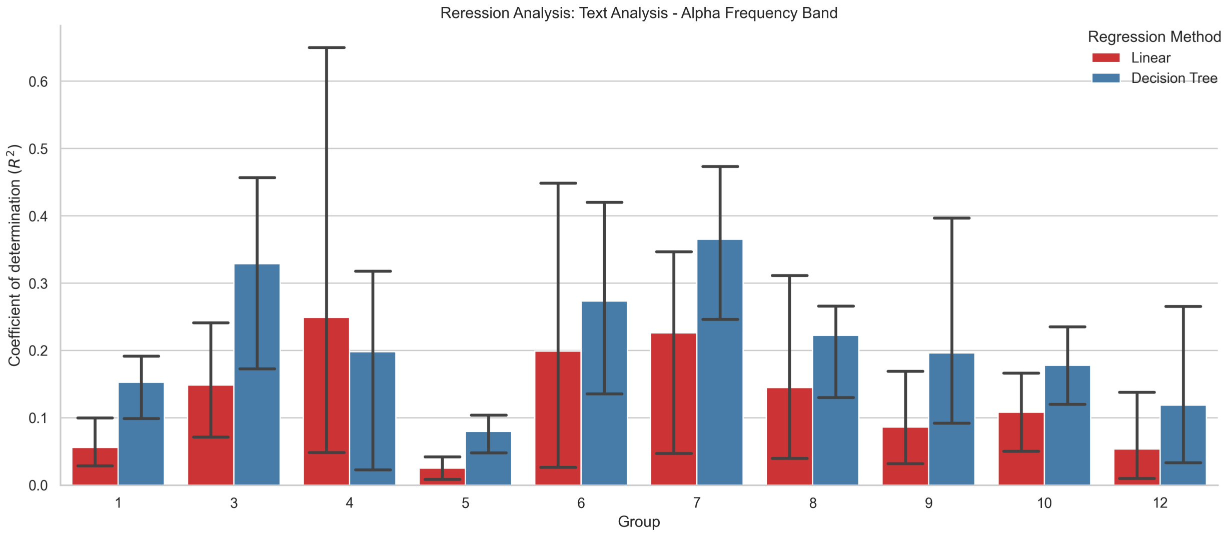


Figure 6.11: Regression modeling alpha band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group.

Table 6.14: Pearson correlation between higher-order team process dimensions and mean group neural measures (outliers removed).

	Transition	**Action**	**Interpersonal**	**Total**
MdRQA	-.21	-0.016	-.64	-.39
MI	-.007	.76	.099	.29

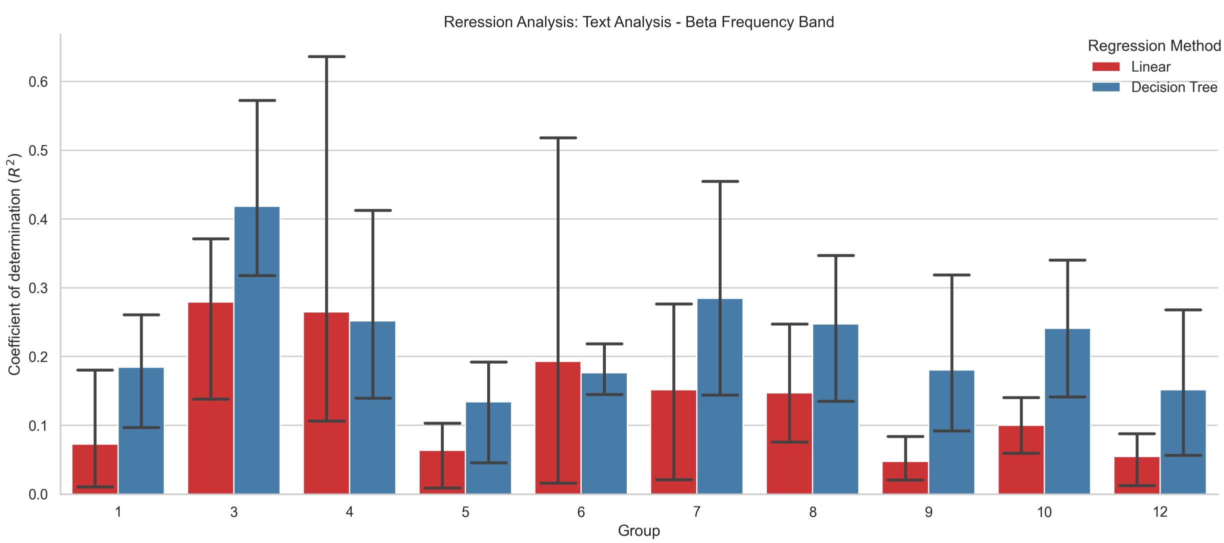


Figure 6.12: Regression modeling beta band power as a function of higher-order team process dimensions for each individual per group. Error bars show the maximum and minimum r^2 values within a group.

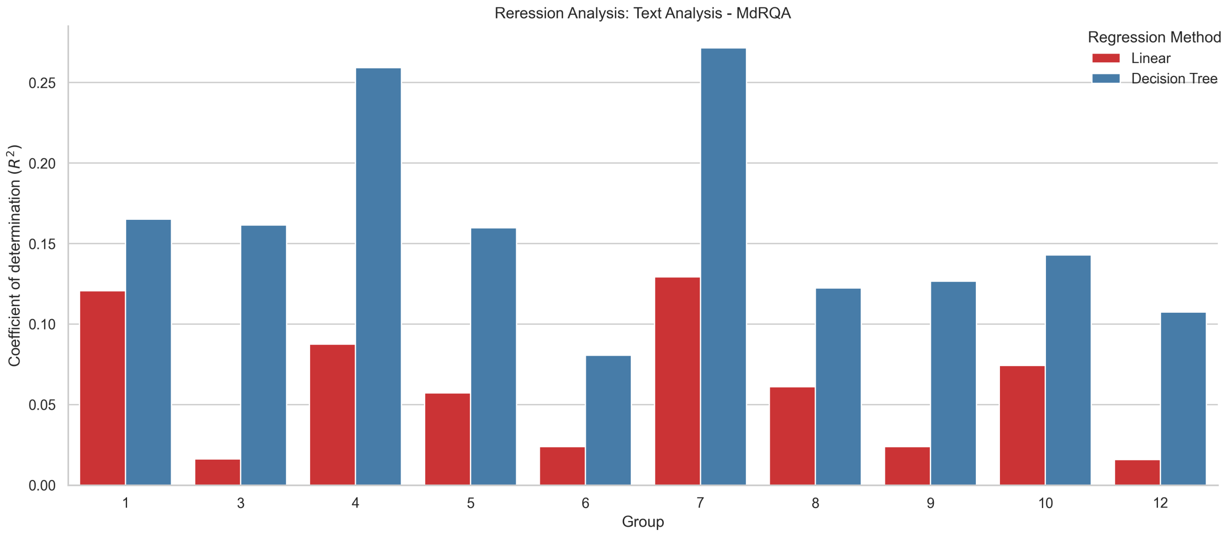


Figure 6.13: Regression modeling MdRQA recurrence rate as a function of higher-order team process dimensions for each individual per group.

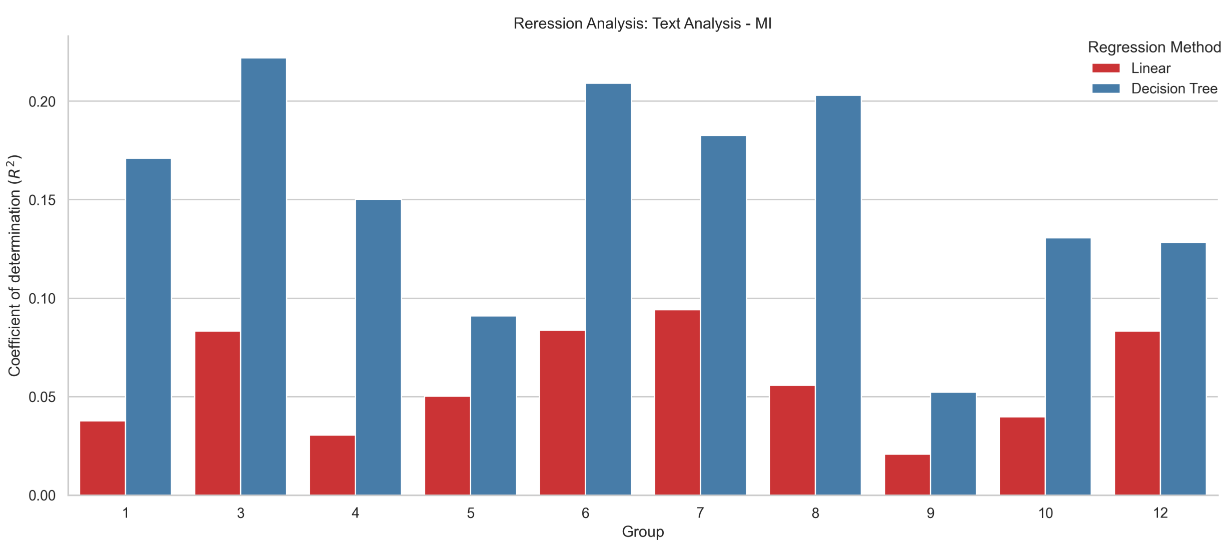


Figure 6.14: Regression modeling mutual information (MI) as a function of higher-order team process dimensions for each individual per group.

6.5 Discussion

Although our analysis is ongoing and largely exploratory, we nonetheless uncovered several promising findings to which we can direct further attention.

Our first research question asked what relationship, if any, exists between neural measures of individual, internal states and neural measures of joint interpersonal states. Of our two individual measures, we found that the task engagement index (TEI) was significantly more closely related to the recurrence rate from MdRQA than mutual information, while the task load index (TLI) exhibited no clear relationship. One of the major distinguishing factors between the two indices is the influence of beta band activity on TEI (TEI = $\frac{\beta}{\alpha+\theta}$ and TLI = $\frac{\theta}{\alpha}$). While the cross-correlation between TEI and recurrence was modest, its existence across participants in several groups suggests that the group recurrence rate reflects patterns of cognitive engagement among group members.

Corroborating the existence of some relationship is the fact that beta activity has been used jointly with recurrence features in prior work classifying EEG signals [253, 372], most directly with recurrence rate for classifying cognitive effort [232]. Despite the existence of this work, we did not find research describing the specific nature of a relationship between recurrence rate and beta activity or recurrence rate and engagement. Future work could explore this further; it remains to be seen whether beta activity drives recurrence, or whether this relationship arises from collaboration or from other factors.

Our second research question focused on examining the relationships between neural measures and behavioral measures during collaboration, including performance and team processes. We generally observed weak to moderate relationships between our variables of interest, though there were some notable exceptions.

There was no clear relationship between prior familiarity with escape rooms and team performance, counter to our initial expectations. We initially assumed team members who had the most expertise—those who had been to many escape rooms, played escape room games, or took classes designing escape room experiences—would be able to provide guidance to the rest of their teammates and ultimately achieve high scores. In actuality, merely having expertise was not enough—in our observation of participants, it became clear that expert team members needed to be *willing* to take a leadership role offering guidance, and their other team members needed to be *receptive* to it, in order for it to be useful. Regression models additionally considering the individual neural indices (TEI and TLI) were modestly better predictors of performance, though still not robust. Considering group measures, mutual information was moderately correlated with team design scores, which could corroborate prior work by Stevens & Gallway [315] relating mutual information to problem-solving during collaboration. However, all of our analyses of performance have low reliability because our predictors were averaged across the duration of the experiment in order to match the number of available performance observations (i.e., one per group), so relevant temporal information is lost.

Arguably more important than performance are the levels of team processes at play during collaboration, which can provide a sense of the overall functioning of the team and its workflow. We did see fairly robust correlations between team performance and levels of team processes: performance negatively correlated with transition processes (goal specification, planning) and interpersonal processes (motivation, conflict and affect management), suggesting that excessive planning or disagreement resolution can be detrimental to performance. After removing data for outlier groups who communicated the least and the most (Groups 3 and 5), action processes (coordinating, monitoring progress) were also positively correlated with performance, which indicates task-focused communication can support higher performance. However, the sensitivity of these relationships to outliers (causing a sharp positive adjustment in the correlation coefficients) supports the need for augmenting our understanding of team processes with data describing *nonverbal* interactions to get a more complete understanding. The fact that Group 3 was one of the highest scoring teams but had the lowest level of process-relevant speech, while Group 5 exhibited the opposite relationship, suggests that other factors in addition to speech are relevant to team performance. We plan to investigate these factors in subsequent analyses.

We observed generally weak relationships between the individual neural indices and levels of team processes across regression and correlation analyses, with the exception of a moderate negative correlation between TLI and transition processes as well as overall process-relevant communication prior to the removal of outliers. It is possible this indicates that engaging in early planning and goal setting results in reduced workload over the course of collaboration, but further data is needed for a robust assertion.

Additionally, theta, alpha, and beta activity were best predicted by different team processes, with alpha and beta most closely related to action processes and theta related to interpersonal processes. The fact that different frequency bands are related to different team processes makes sense because of their association with different cognitive functions—beta activity is associated with task-related cognitive processing; alpha is associated with relaxation, and inversely associated with sustained attention and engagement; and theta is associated with relaxation or emotional processing [121]. Considering interpersonal processes such as affect management or conflict management might require emotional processing, and action processes such as coordination or systems monitoring might require changes in attention and cognitive effort, one would expect an association with the frequency bands indicative of these necessary cognitive and emotional processes. The fact that the raw frequency bands were more related to team processes than the composite TEI and TEI measures, which contain contributions from multiple frequency bands, may be interesting to explore further.

Finally, although the group neural measures did not have strong relationships with team processes when all group data was considered in regression and correlation analyses, repeating the correlation analysis between the group mean the neural measures and the total levels of each team process for per group—and thus eliminating the temporal characteristics of the data—yielded our most robust relationships, which were further bolstered upon the removal of outliers. Mutual information was strongly correlated with action processes, supporting prior work from Stevens & Gallway [315], who observed increased mutual information and decreased symbolic entropy in team member EEG recordings during periods of problem-solving. Similarly, recurrence rate was strongly negatively correlated with interpersonal processes, mirroring findings from Eloy et al. [85], who found a negative relationship between recurrence rate and all team process dimensions in fNIRS recordings of dyads performing a collaborative task. While the negative association may seem surprising, Eloy et al. report that such decreased regularity in physiological and behavioral signals can yield increases in performance and other facets of collaboration. Teams that revisit or remain in the same state, and get "stuck," while teams with more irregular dynamics are more adaptive. In our study, these more adaptive teams may have spent less time resolving conflicts about aspects of their designs, reflecting their levels of interpersonal processes. The fact that our data can replicate these prior findings indicates that we are able to detect collective brain dynamics relevant to collaboration in the context of our experiment, which shows promise for the development of a BCI system able to leverage these processes.

This result is also interesting because it arises from an average across time, much like event-related potentials do in EEG recordings from event-driven studies. Where other similar studies typically measure team processes or other measures of collaboration quality via self-report surveys after the task, we took care to preserve as much temporal information as possible in our study design and recording setup; the ability to respond in real-time to new information might be a necessary capability for a future BCI support system for teams. Since a clear relationship exists between group neural measures and team processes in the aggregate data, but not at the temporal resolution of our time series, it is likely that we will need to use a larger window size when calculating the group neural measures and levels of team processes in order to accurately capture these dynamics. We will explore this in future work.

6.5.1 Limitations and Future Work

The analyses described in this chapter represent first steps toward exploring the relationships between the brain and behavioral dynamics of team members during collaboration, and yielded some promising results. Nonetheless, there are still many more aspects of our data we wish to explore in future work. First among these are the two remaining group neural measures, the driver-empath model and wavelet transform coherence, which we were unable to include in our analyses due to time constraints, but would provide another dimension with which to examine how group neural dynamics relate to team behavior. We also plan to incorporate analysis of responses to the PMERQ to assess how the emotion regulation strategies of team members impacts team performance. Following our discussion above, we also plan to redo our analyses and compare all measures using different window sizes, to see whether stronger relationships between team processes and the different neural indices emerge.

While our analysis of team processes is theoretically grounded, our approach is limited due to the nature of our task, where a large amount of collaboration between team members happens non-verbally, either with one another through interaction with the Miro digital whiteboard, or mediated through interaction with ChatGPT. Measuring team processes purely through the text from transcripts is inadequate for describing the full breadth of the interaction context. Drawing from prior work, there are several possible approaches we can take to capture more information about the non-verbal dimensions of participant interactions. One straightforward first step is to categorize participant interactions while they used Miro to document their designs. Adapting the approach from Aytes [15] since it is applicable to our data, we can classify participant behavior as either *parallel* (each team member working separately), *interactive* (all team members working together on the a single design element), or *scribe* (one team member is chosen to document the group's ideas). It is also possible to log the object manipulations participants made (i.e., when objects were created, destroyed, and modified). We can also further investigate how participants used the AI. This could merely be frequency of use [301], or we could develop a coding scheme to categorize the types of interactions people had [124]—from our observations of participants, these might be, e.g., asking for ideas for escape room themes, or asking for ideas about potential puzzles. Finally, we can use participants' camera recordings to collect data about their body posture, facial expression, or eye gaze over the course of the experiment [104].

Furthermore, although adequate for initial exploration, our regression analyses are limited because they only model the data of individuals or groups, rather than taking into account the whole of the data. It would be more appropriate to model team processes using a mixed-effects model with participant group as a random effect to account for between-group differences. This could be further extended to a multi-level model to account for effects at the individual, group, and population levels. It would also be interesting to explore the reverse direction of our regression models in this study to model team processes as a function of different neural measures, rather than the other way around.

Finally, once all the data for this study is analyzed and relationships between all variables fully established, we plan to conduct a follow-up study where we provide feedback to participants derived from the brain data that was collected during collaboration, in order to determine how best to present such feedback to users. We would take note of factors such as whether users are able to distinguish between others' feedback and their own, whether they would change or alter their behavior in response to this feedback, and what narratives they form about their experiences over the course of interaction. Eventually, we hope to be able to test a fully operational prototype.

6.6 Conclusion

This chapter examined the extent to which neural and behavioral signals can reflect the quality and nature of collaboration in small teams engaged in a creative design task. By analyzing EEG-derived metrics of individual cognitive states and group-level synchrony alongside detailed observations of team behavior, we identified modest but interpretable links between neural activity, communication patterns, and collaborative performance. Recurrence and mutual information measures showed the strongest associations with task-relevant behaviors, while cognitive load and engagement were modestly predictive of team outcomes. Though exploratory in nature, these findings provide early evidence that brain-derived signals can offer meaningful insights into team dynamics. This work serves as a step toward identifying which types of neural information might be useful for future support systems designed to enhance team collaboration.

Chapter 7

BCIs for Teamwork: Exploring the Potential for Brain-Computer Interfaces in Collaborative Contexts

Abstract

Engaging in collaboration can enable greater efficiency and output than working individually. However, teams are also vulnerable to process loss, where suboptimal working environments, member decisions, or group dynamics lead to inefficiencies that prevent them from reaching their full potential. Recent work has shown that brain signals recorded via hyperscanning can be used to reliably and non-invasively detect levels of individual and team processes indicative of whether a team is functioning effectively. However, to date there are no existing interventions that leverage this ability to facilitate effective teamwork in collaborative contexts. To explore the needs and concerns of the various stakeholders that might benefit from such an intervention, we employed a user-centered approach by engaging with members and leaders of various teams, and synthesize insights for future designers and developers.

7.1 Introduction

Prior work has shown that collaboration between individuals has the potential to fuel creative synergy, where new cognitive inputs, the combinations of personality characteristics, or interaction dynamics can yield a volume, breadth, and fluency of creativity greater than what would be possible from working alone [191]. However, an increase in creative output as a result of collaboration is not guaranteed; indeed, several studies have shown that creative collaboration can result in *decreased* output, despite the potential gains. Such process loss can be caused by a number of mechanisms, including groupthink (the tendency of group members to adopt the majority perspective), social anxiety or apprehension toward sharing ideas, downward comparison with others, cognitive load, or distraction by other group members [191, 85, 68, 258, 261]. Thus, while it is possible to have high-performing creative teams, the appropriate environment, resources, and workflow must be fostered for creative synergy to occur.

Monitoring and supporting effective cognitive states and team processes can enable support systems to maximize team potential and mitigate process loss. Such systems work best when they are unobtrusive and do not impair the team processes of team members, instead monitoring team dynamics in real-time to facilitate the social and cognitive processes necessary for successful collaboration (coordinating, persuading, planning, etc.) [325, 260]. In addition to containing information about relevant individual cognitive states such as mind wandering [202] and workload [13], recent work has shown that brain signals recorded using functional near-infrared spectroscopy (fNIRS) can be used to reliably detect levels of these team processes in team members engaged in collaborative problem-solving [85]. fNIRS recording is non-invasive and robust to motion-induced artifacts, making it useful as input for a brain-computer interface (BCI) that could serve as a creative support system.

While some existing systems from prior work have been introduced that use BCIs to enable creative output or provide support for design and ideation [328, 32, 47, 141, 371], the vast majority have focused on assisting *individual* users. Very little work has been done employing brain sensing specifically in the context of *group collaboration*, and even less has employed *hyperscanning* to leverage data from multiple users simultaneously (see [235] for one example). To date, no general-purpose BCI support tools for collaboration exist that leverage the relevant phenomena they can sense to assist real-world users, and the perspectives of relevant stakeholders have not been assessed to inform the development of one. With the increasing rate of research on brain-computer interfaces in the real-world [275, 310] as well as the commercialization of wearable brain sensing devices (e.g. Neurable MW75 Neuro, Mendi, Cognixion Axon-R, OpenBCI Varjo, etc.), it may be possible for team brain-computer interface tools to become a reality in the near future.

However, proceeding directly with the design and implementation of a system incorporating user physiological data has several potential risks. There is a chance of creating a system that users do not find useful, and thus will not want to use; or that the system would rely on measuring or displaying physiological markers that users may not feel comfortable sharing with others, such as their level of engagement or workload during team meetings. It is possible to develop a system both team members and leaders find useful, but has potentially harmful downstream consequences, such as enabling new forms of discrimination based on brain activity, or allowing employers to tie cognitive metrics to employee performance reviews. In order to mitigate these issues and concerns, it is necessary to integrate user perspectives in the earliest stages of designing the system, before any prototyping takes place.

Therefore, in this chapter we outline an exploratory study evaluating the needs and concerns of stakeholder groups that would be impacted by the introduction of a BCI support system for collaboration, seeking answers to the following research questions:

1. What unmet (or under-met) needs do stakeholders (team members and team leaders) have over the course of collaborating?
2. In what ways could a BCI help fulfill these needs?
3. What concerns do stakeholders have about a BCI being used for this purpose, and what risks do they foresee?
4. How could such risks and adverse impacts be minimized over the course of design and development?

Specifically, we adapt user-centered approaches employed by Holstein et al. [133] to assess user needs and develop design requirements for emerging technologies, entailing a series of design interviews and brainstorming exercises with the members and leaders of various teams.

7.2 Exploring User Needs: Design Interviews with Team Members and Leaders

In order to examine stakeholder needs and preferences for working on teams, as well as regarding the use of technology to assist during collaboration, we employed a user-centered approach utilizing several different design interviews. Here we outline our study design and summarize key findings.

7.2.1 Participants

We recruited 11 participants (7 female) that were either members or leaders of teams engaging in creative collaboration. To avoid unnecessarily restricting perspectives that may ultimately prove insightful, we use the definition of creativity from Amabile & Pratt [8]: "the production of novel and useful ideas by an individual or small group of individuals working together." This is contrasted with *innovation*, which is "the successful implementation of creative ideas within an organization." Thus, for the purposes of our study, a team engaging in creative collaboration is one engaging in any form of creative problem solving; our cohort included educators, software engineers, a marketing professional, healthcare professionals, student researchers, a university sports coordinator, and more, several of whom also engaged in team activities as hobbies. Participant backgrounds are provided in Table 7.1. Recruitment was conducted via email outreach to potentially eligible participants.

Table 7.1: Study participants and their respective team backgrounds. Participants marked with an asterisk (*) were unavailable for the second experiment session.

Participant	Background
P1	Graduate student mentoring undergraduate project teams; former officer on the executive board of university's graduate student robotics honor society
P2	Former release engineer and program manager for a well-known online retail company
P3*	Chief operating officer of a nonprofit supporting technology companies and the tech ecosystem more broadly in Massachusetts
P4	School psychologist in the special education department at a large regional high school
P5	Former estimator/current project engineer for general contracting company overseeing construction projects in a large metropolitan area; player in a recreational baseball league; singer in a church choir
P6	Part of a team of educators at a university-based center providing professional development for Pre-K–12 teachers as well as engaging in education research community outreach
P7	Veteran service representative for the Department of Veterans Affairs; Dungeons & Dragons party member/Dungeon Master with a group of friends
P8*	Team leader in account management for an advertising technology company
P9	Graduate student mentoring undergraduate project teams, and leader of a student AR application development team
P10	Director of physical education and athletics at a university
P11*	Team member at a concierge health services provider

7.2.2 Study Design

Adapting the design of Holstein et al. [133], the study consists of three different types of design interviews, conducted remotely over the course of two Zoom sessions. The procedure for each session is outlined below.

The first Zoom session consisted of:

1. **Generative card sorting with "superpowers" (Part 1):** Participants were asked to consider their roles on any creative teams that they typically engage with, and to think about and describe the sorts of challenges they face. They were then asked, "If you could have any superpowers you wanted, to help you accomplish your goals when working as part of these teams, what would they be?" Participants were instructed to create sticky notes with their desired superpowers using Miro, an online platform for visual collaboration, and then sort them based on their relative priority. If this process inspired new ideas, participants were encouraged to add new notes.

2. **Semi-structured interviews:** After ranking their superpowers, we conducted semi-structured interviews with participants in order to better assess their needs for support and to investigate the role supportive technology could play in providing this support. Participants were asked to reflect on their experiences on the teams they were a part of, any challenges they have faced, whether they had used any technologies to overcome these challenges, and to imagine how a technology that could detect how well teams were working together could improve their team interactions, assuming it had no limitations. We also asked participants whether they had experience using any generative AI tools, and what role if any they play in the workflow for their teams.

After responses from the first session were compiled and synthesized, the final design interview took place in a second Zoom session. Of the 11 participants who completed the first session, 8 (4 female) were available to participate in the second. The components of the second session were as follows:

3. **Generative card-sorting with "superpowers" (Part 2):** Following the initial sorting, participants were presented with a representative selection of cards based on responses from the first session (see Section 7.2.3) and asked to rank them as well. If any of these new superpowers were considered equally important or redundant with other powers that had been generated, participants were encouraged to align them vertically to indicate a tie. If the participant did not desire one or more of the new superpowers, they could be omitted from the ranking.

Once the study was complete, the relative rankings of the superpowers were compared between each of the participants to explore trends.

Table 7.2: Results from the superpower card-sorting exercise from the first experiment session. Shown are the superpowers which were generated by multiple participants, and were chosen as most important for at least one. The total frequency of occurrence across all participants and the frequency of selection as the most important power are provided in the second and third columns, respectively.

Superpower	**# Total Occurrences**	**# Most Important**
Slow down time/ infinite time	6	3
Mind reading	5	2
Build trust/ change attitudes	5	2
Better communication/ understand perspectives	6	1
Cloning	3	1
Super intelligence/ memory	3	1

4. **Speed-dating possible futures:** To explore potential concerns and design tensions arising from the prior session and further probe and evaluate user needs, we presented storyboards of futuristic scenarios based on the ideas presented during card sorting and semi-structured interviews. These scenarios were shown to participants in quick succession in "speed dating," a design method for rapidly exploring new technology concepts. The scenarios were designed to probe potential concerns or ethical issues that could arise if a BCI technology for supporting creative collaboration was implemented based on ideas and needs raised in the earlier session. Further details are provided in Section 7.2.5.

7.2.3 Generative Card-Sorting with "Superpowers" to Probe Challenges Faced by Teams

Initial card-sorting results from the first experiment session are summarized in Table 7.2. In general, participants desired a way to ensure and keep track of the level of shared understanding between team members, as well as ways to motivate and engage the team members they worked with. Some also focused on the team's objective performance, indicating they wanted to increase their own performance or that of team members.

To ensure that common trends among the superpowers participants generated in the first session were synthesized into a tractable number of representative powers for the second session while preserving the breadth of responses, we engaged in a collaborative clustering approach. First, the we consolidated similar and duplicate superpowers, then generated further superpowers to clarify participants' intentions for their chosen superpowers based on their thinking aloud during the session (i.e., one participant might have chosen "mind reading" because they wanted to understand the thoughts and intentions of teammates, while another might have chosen it to know whether misunderstanding or miscommunication had occurred). Finally, we grouped the new reduced set of superpowers into representative groups based on their similarity, iterating until agreement was reached on the number of groups and their composition. Clearly descriptive names for each cluster of superpowers were chosen as new representative superpowers to use for the second session. In total, 16 representative superpowers were generated: "Align team effort and output," "Align team motivation and attitudes," "Knowledge of what team members are thinking/feeling," "Control over external factors that can limit the team," "Boost your own skills/abilities," "Eliminate travel time," "Build and maintain relationships with clients/others not on the team," "Avoid errors," "Accomplish more in less time," "Ability to reduce stress and burnout," "Understand team assets and skills," "Build and maintain trust/camaraderie among the team," "Know how the team can be helped and improved," "Build confidence," "Transfer knowledge, skills, and understanding to others," and "Detect and avoid miscommunication/misunderstanding."

Results from the second session are summarized in Figure 7.1, which depicts aggregate pairwise comparisons of the relative importance of the superpowers ranked by participants. Note that two participants (P5 and P7) completed two different rankings, each with respect to a different team they were part of, so the maximum possible agreement is 10. Participant rankings for this portion of the experiment were rather heterogeneous—some participants ranked all 16 superpowers linearly with little or no ties, some grouped powers into several columns of one to five powers of equivalent importance, and others chose to make three or four large sets of groupings; one participant discarded several powers as not applicable or unimportant, while most others did not discard any. Nonetheless, when viewed as a whole, it is possible to draw some conclusions about their preferences. In general, the most important superpowers were those that would enhance knowledge about *how a team is working* and *how it could be improved* (e.g., "Know how the team could be helped and improved"; "Detect and avoid miscommunication/misunderstanding"; "Transfer knowledge, skills, and understanding to others"), versus changing specific qualities about the work and work environment ("Control over external factors that can limit the team"; "Eliminate travel time"; "Build and maintain relationships with clients/others not on the team") or the team ("Align team effort and output"; "Align team motivation and attitudes"; "Build confidence").

An interesting phenomenon to note is that the "Ability to avoid errors" was generally seen as important, but less important than “Boost your own skills/abilities” (which was considered of medium importance) compared to preferences for other similarly prioritized superpowers. Participants who prioritized the ability to avoid errors over boosting their own skills (P2, P7, P9, P10) tended to emphasize the importance of deliverables and favorable performance in their line of work, or cared more about the cohesiveness of the team and productive relationships between team members versus their own individual abilities. Participants who viewed avoiding errors as less important tended to rank it at or close to the lowest priority; of these, several noted that making errors could actually be beneficial, because it could allow team members to learn and grow from their mistakes.

Finally, while it tended not to be the most important consideration, the “Ability to reduce stress and burnout” was a priority for many participants. Some participants reported feeling burnt out from their work, and several noted that stress and burnout could result in several other negative downstream effects for the team, such as reduced motivation and ultimately reduced productivity.

7.2.4 Semi-Structured Interviews: What Works Well, What Doesn’t, and How Tech Could Help

We analyzed audio transcripts from interviews using two techniques from Contextual Design [134]. First, we examined the transcripts through the lens of interpretation sessions, through which we extracted quotes representing key issues and insights. We then constructed affinity diagrams using the extracted quotes. An affinity diagram is a multi-level hierarchical clustering method whereby higher-level categories gradually emerge from bottom-up clustering. First, quotes were grouped into 329 unnamed level-1 categories based on their perceived level of similarity, and then labeled. These categories were then grouped into 29 level-2 categories, and the process was repeated until categories representing more abstract, high-level themes were reached. In total, six top-level categories emerged from the 29 mid-level categories:

1. Team member differences can be both an asset to use and a challenge to overcome

2. Clear communication is crucial

3. There are benefits and drawbacks both to working in person and remotely; a workflow that leverages both approaches might be best

4. Tech can be used to inform team decision making and alter or augment team workflows to provide interventions to make them more effective, minimizing burnout and process loss and maintaining team morale

5. Desire for knowledge about team members’ mental states, their dynamics with one another, and contributions to the team

Key: Cells contain number of participants who ranked row superpower as more important than column superpower	Know how the team can be helped and improved	Detect and avoid miscommunication/misunderstanding	Avoid errors	Transfer knowledge, skills, and understanding to others	Knowledge of what team members are thinking/feeling	Ability to reduce stress and burnout	Build and maintain trust/camaraderie among the team	Boost your own skills/abilities	Understand team assets and skills	Align team effort and output	Align team motivation and attitudes	Build confidence	Accomplish more in less time	Control over external factors that can limit the team	Build and maintain relationships with clients/others not on the team	Eliminate travel time
Know how the team can be helped and improved	0	4	5	7	6	7	5	6	6	5	6	7	8	7	9	9
Detect and avoid miscommunication/misunderstanding	5	0	6	6	7	6	4	7	5	6	6	5	7	7	8	9
Avoid errors	3	3	0	4	4	5	5	5	6	6	7	6	5	7	7	8
Transfer knowledge, skills, and understanding to others	2	2	4	0	3	5	4	8	4	4	5	5	7	6	7	10
Knowledge of what team members are thinking/feeling	1	2	6	3	0	5	3	7	3	3	5	5	7	5	8	9
Ability to reduce stress and burnout	2	2	5	3	5	0	3	7	3	4	5	5	7	4	7	8
Build and maintain trust/camaraderie among the team	2	2	5	4	4	5	0	5	2	3	4	5	6	5	6	8
Boost your own skills/abilities	2	3	3	2	3	3	5	0	4	4	5	5	4	6	7	8
Understand team assets and skills	2	2	4	4	4	5	2	5	0	2	4	4	6	5	7	8
Align team effort and output	3	2	4	4	4	4	3	6	3	0	2	5	5	4	6	8
Align team motivation and attitudes	3	1	3	3	4	4	4	5	3	2	0	3	5	6	5	7
Build confidence	3	2	4	3	5	2	2	3	3	3	5	0	4	5	6	6
Accomplish more in less time	2	2	4	2	3	0	3	3	3	4	5	3	0	4	7	7
Control over external factors that can limit the team	1	1	3	2	5	2	3	4	2	3	2	3	4	0	6	8
Build and maintain relationships with clients/others not on the team	1	0	3	2	1	1	1	3	1	1	2	1	3	3	0	7
Eliminate travel time	1	0	2	0	1	0	1	2	1	1	2	2	3	2	2	0

Figure 7.1: Pairwise comparison of rankings for each superpower in the second experiment session. Numbers and colors in each cell indicate the number of participants who ranked the superpower in the corresponding row greater than that in the corresponding column. Higher numbers/darker shades indicate greater agreement between participants. Note that two participants (P5 and P7) completed two different rankings, each with respect to a different team they were part of, so the maximum possible agreement is 10. Powers are sorted top to bottom in order of participant agreement on their relative importance (i.e., the sum of the respective rows).

6. Concerns about privacy, equity, and usability of technology to help teams

Team member differences can be both an asset to use and a challenge to overcome

One of the most common challenges participants encountered on their respective teams was friction between team members due to differences in motivation or incompatible personalities, a phenomenon consistent with prior work. In such cases, team members not pulling their weight and performing the duties asked of them, or who were otherwise resistant to necessary changes to their typical workflows (such as using new software tools or executing new company policies) led to decreased performance, frustration, and failure to meet deliverables. As described by P2,

> "One of the biggest challenges is always...having people come to the meeting, come to the table very opinionated, set in their ways...that don't want to listen to new ideas. Or accept change. Even when change means growth. You know, people get comfortable...people get comfortable in what they're doing. And it's working for them. So, making a change is scary."

Similarly describing their experiences with former team members one the executive board of their university's graduate student robotics honor society, P1 recalled

> "The people who were in the past involved really did not mesh well with my personality and work style. And so I'd get really frustrated. And so I think like. A challenge is personalities. Personality mismatch that just doesn't feel like you mesh, feels like you're oil and water with somebody."

Despite the challenges that arose from these incompatibilities, however, one of the positive aspects of working on teams that participants cited most frequently was the ability to learn from and take advantage of the diverse backgrounds, expertise, and collaboration styles of other team members, ultimately forming wholes greater than the sum of their parts. According to P6,

> "I think personalities are really the big thing, and having different personalities and different perspectives. And while that's really challenging, recognizing like, the importance of that, that everyone brings different perspectives and that makes like whatever your end product is better and stronger."

Similarly, P8 describes how their professional experience was enriched because of their colleagues' different skillsets and areas of expertise:

> "I think just being able to learn from and take advantage of different people's skill sets is so valuable. Like for example, like I know, people who work in sales who are, you know, just like extremely well connected. They have a really strong understanding of the politics of different companies that we work with. They're particularly good at client facing skills. I know I mentioned also I work with like a trading team that's more, in the platform every day and, you know, they tend to have like a deeper level of knowledge about the product and more technical savvy and I think I learned so much from working with those different teams also even just within my own account management team. Obviously people have different backgrounds and they might have worked on different kinds of accounts throughout the years. So everyone brings something unique to the table."

Clear communication is crucial

Another common sentiment participants expressed during their interviews was the overwhelming need for clear communication between team leaders and other team members, with several relating experiences when team performance suffered because of miscommunication or misunderstanding. As P7 summarizes,

> "Let's say the person who explained ABC isn't fully clear and the person that's really supposed to be A thinks they're supposed to do B. And then the person who's supposed to do C is doing A. If you don't clearly communicate that, and everyone is doing their own little bit of things, you know, A B and C aren't actually really done. So to speak. And then the whole team would suffer for it."

Many other participants expressed similar frustrations regarding occasions where important work was not adequately completed because of an incorrect understanding of the task at hand; P8 likewise notes

> "Something that I've seen at my current job and responsibilities and even at my previous job is like people either accidentally stepping on each other's toes or maybe not doing the work that other people expected of them. Because for whatever reason, there was some confusion about who was responsible for a specific task."

Thus it is extremely important that those with leadership roles on teams are able to ensure their directions or guidance are understood by those reporting to them, though this is far from easy. Recalling their role mentoring an undergraduate student project team, P1 stated

> "I can give all of the energy and support possible, and they could still not understand what I'm trying to say and the project could still not turn out to be successful no matter how much effort I put in as a guide."

Both in explaining their reasoning for the superpowers they chose as well as when describing ways they wished technology would assist them on their teams, participants consequently expressed desiring ways to mitigate such miscommunication or misunderstanding. When stating why they chose “mind reading” as their most important superpower, P3 explained

> “As a manager, I would use the mind reading to kind of know what the team is thinking, right? So sometimes, I don’t know if you’ve experienced this in meetings, sometimes people will nod and smile and then, you know, 3 hours later it’ll be very clear that they didn’t understand and didn’t ask the questions. And so, if I could read their minds, then I could, you know, sort of address issues and questions in the moment.”

Describing a technology they wish existed to help presenters ensure that their fellow team members were engaged audience members, and that key takeaways were understood and delivered convincingly, P2 stated

> “I know that there’s reactions they can have to what’s being presented to them, what’s being told to them. It would be, you know, and again, if you’re in a meeting and you’re talking to 40 people, you’re not gonna pick up on all of that for all 40 people. You might notice it in one or two of them at best. So if there was, some sort of the technology that was scanning, looking at the people and interpreting the way that they’re holding themselves to present like,...a little dashboard that could give you an indication, if I was talking to 50 people and 47 out of 50 were green and two were yellow and one was red, I’d be like, “Okay, that’s good, I read the room well, I did a good job, I gave them what they wanted, and they believed me and accepted what I told them.” Whereas if I get, you know, 25 reds because you know people are unhappy with what I’m saying, I’ll be like, “Okay, that one didn’t go well.”

Considering these responses from participants, a tool to reduce or eliminate this sort of miscommunication would likely be considered useful to others working on teams.

There are benefits and drawbacks both to working in person and remotely; a workflow that leverages both approaches might be best

A large majority of participants in our cohort worked on teams which allowed hybrid work; only P5 worked only in person, and only P11 worked entirely remotely. While this tended to afford participants with a great deal of flexibility in executing their responsibilities, many also noted benefits and drawbacks in equal measure to both remote and in-person work, favoring each approach for different reasons. Working remotely, for instance, eliminated the time required for commuting, and in some cases allowed face-to-face meetings to occur where distance would have otherwise rendered them impossible, or very infrequent, as noted by P2:

> "There's a lot of meetings where there's people from multiple locations and it doesn't really make sense to think that we would ever all sit around the same table except for very occasional times. So, having the ability to do something like this, certainly helpful for being able to quickly jump together and be able to see somebody face-to-face and and talk to them and so, that's a helpful thing to hold the concept of, Zoom or any other meeting, video meeting capability."

A remote digital workflow can also eliminate the need for physical space and materials; P9's AR application project team initially started their work in-person and used a physical kanban board (implemented with sticky notes on a whiteboard) to keep track of task progress, but eventually moved to a hybrid workflow and switched to a digital alternative, which was easier to modify and share remotely.

Though it can afford convenience, a potential downside to remote work that was reported by participants was the increased difficulty in "disconnecting" from their digital environment. Several participants described relying on asynchronous messaging applications to keep in contact with their fellow team members, and the need to implement strategies to use these tools effectively while mitigating distractions, as exemplified by P6

> "I have like my work Outlook on on my phone, but I don't, I don't have like notifications popping up that I got emails or anything, like I turn those notifications off. I really struggled with even just like Slack, 'cause I felt like that was constantly notifying me if things I didn't necessarily need to know about."

The need for a stable internet connection for video calls was indicated as a possible challenge as well.

Alternatively, some problems may be easier to solve in person. As P11 notes,

> "I know I said I like working remote, it's great. I love it, but I do think that that partially comes with a challenge because you don't have someone to just like pop over and ask a question to. You have to either go to email, phone or, or Zoom, but like it's not as easy. Like when I started my other job, he would just literally be like, "oh, so like, hey, what, what do you do for this?" You can't do that as easily remotely."

P10 likewise indicated that in-person conversations were sometimes more effective:

> "I can do videos in my role to alumni and and others and you know being interviewed and all that. So I'm fine with that but as far as the form of communication, there's nothing more powerful than walking into somebody's office, closing the door and sitting down and having a conversation."

Technology can be used to inform team decision making and alter or augment team workflows to provide interventions to make them more effective, minimizing burnout and process loss and maintaining team morale.

All participants in our cohort used some form of technology to streamline some aspects of their day-to-day workflows and improve their team's organization and efficiency. Due to the diverse nature of the teams represented, the specific tools and platforms used by each participant tended to differ, though there was a degree of overlap with commonly used consumer software. In particular, all participants regularly used some form of video conferencing software (typically Zoom or Microsoft Teams) for remote meetings, as well as tools for asynchronous communication (sometimes email and text messaging, but increasingly instant messaging platforms such as Slack or Teams). Several participants also used cloud-based document collaboration platforms, such as Google Docs Editors, Microsoft Office, or Confluence.

Other software tended to be tailored for executing the specific job duties of P3, P4, P5, P7, and P8 (e.g., construction planning/estimation, filing benefits claims, etc.). Another relatively common use for technology platforms by some participants was to organize and allocate tasks, and keep a record of what has been done and what is left to do.

P1, for example, used Canvas (typically a learning management system for students) for this purpose:

> "For the [student project], I use Canvas to assign the team... I guess assignments, that are trying to like, help push them along or like think more critically about a certain topic and that has been working really well."

P6 similarly describes their experience using Milanote, an organizational tool for creative projects:

> "We're using Milanote as a tool. That's one of our favorite tools to use. And it's helpful. I like that I can go back to see what we were talking about. Maybe it's like notes from a meeting or things [my supervisor] is thinking about, or maybe it's like a checklist of things that has to happen. So, it's helpful to have like documentation of what we did."

Participants agreed that the tech tools they regularly used mitigated some cognitive overhead and helped them be more effective on their teams. Nonetheless, when asked about technologies they wished existed to help them on their teams, they identified many aspects of their work and interactions with other team members which could still be improved, and were optimistic that future technologies could be used to make their teams more effective.

Several ideas participants raised centered on detecting if the team was veering off course or not operating to its fullest potential, and providing some intervention to get people back on track. As suggested by P3,

> “If a tool could detect where things are missing and help you prioritize, that would be the other thing that a tool like this could do is, there’s deadlines that I miss or there’s projects that I miss or that I don’t get to... but are they the important ones? Or am I how am I spending my time doing the things that aren’t priority that aren’t moving me forward just because they’re tasks that I can spin on. So a tool that could help identify, but would have to get into somebody’s head somehow to. Or know or know the keystrokes or know that I don’t know what I don’t know how it would do it but yes that would be amazing. To kind of give you a sense of where things are getting. Getting stuck because of miscommunication or misunderstanding or other blockers.”

P1 suggested a system that could help resolve impasses or conflicts between team members, or direct them to seek out more targeted assistance:

> “I think if there was a way that it could either like, send you to a place where the issues could be solved. Or have some sort of like not escalating way to discuss the problem. Like if it sends me a message like, “hey, by the way, this team member like gets really annoyed when you write the to-do list because she’s the secretary and she should be writing the to-do list but she’s not, so that’s why you do it.” Even a messaging system instead of maybe direct confrontation just being like “So-and-so is like not feeling really great right now.””

Others suggested a system to help with managing employee workload, so team leaders could more effectively balance tasks and maintain morale:

“I think I would like it to tell me... How other people are responding to their current workloads, and what each individual person’s, I don’t know, stress level is so that we can come together more as a team to, you know to problem-solve the pervasive issues that are going on in our department.” — P4

“I’d like to know if they’re burnt out, if they feel like they’ve given their all and they need a break or something. If they feel insecure about, if they feel vulnerable, knowing the morale of the team would help me know how to manage it, I guess.” — P5

Finally, others noted that a system able to automatically provide positive feedback for teams which performed well would benefit team member well-being. For example, P1 stated

> “If we were doing good, then I think it would be something like, “Oh, look at how great we are.” It’s like in addition to like the feeling of like this is a great team, there’s like a satisfaction of like some sort of score or number or green light.”

P6 also advocated for a system providing positive reinforcement for teams that were performing well:

> “I really like thinking about the positive, positively reinforcing. Sometimes I would ask my groups, I would like tell them that I want to hear you like praising people, other people’s ideas, like maybe we were brainstorming something and I say I purposefully want to hear, “I’m gonna walk around and listen for...” and I would ask them to like, so I would like tell them, “I want to hear praising.” And so it would actually change the dynamic of the group because they would be wanting to get, I don’t know if I did bonus points or whatever, but like you know, they would get that reaction from me like, “Nice job.” So I think the positive interactions would be good.”

Desire for knowledge about team members’ mental states, their dynamics with one another, and contributions to the team

Related to other suggestions for possible technologies enhancing collaboration by providing specific interventions, participants also expressed interest in technologies which would measure attributes relevant to teamwork, such as the relative contributions of each team member, their mental states, and the team’s collective dynamics. This approach could afford teams flexibility to tailor responses to the measured metrics to their specific needs, or allow them to examine the relationship between these variables and the quality of team members’ work or day-to-day experiences.

P10 was strongly in favor of collecting data to measure team members’ level of contribution to their teams, since it would improve accountability:

> “Imagine if you’re in and it’s not with an athletic team like, you know, we get management council together, right? And we wired it somehow that you’re able to see. And be able to provide feedback individually of like, are you bringing your best attitude today? Are you really paying attention? Are you really contributing? Are you really listening? Like, to be able to measure that in a team? To be able to give that feedback that people like, look, you’re not holding up your end of the bargain in this teamwork. That would be, I think, revolutionary.”

Participants also proposed systems which would measure the quality of relationships between each member of the team, both to provide incentive for having productive and positive interactions but also to examine and address any tension or lack of synergy that might be observed. P1 outlines what they imagine what a new technology supporting collaboration might be like, stating

> “If we had a technology like this, I feel like I would be able to see like my, attitude or disposition with each person, and with each other between the team. And I feel like I can picture like, green light green light green light and then like orange

> light, kind of...[I envision it being like a map], where I can see the connections between the other people are good. And even if that's like a map of like lines connecting the nodes of people."

P6 had a similar idea, modeled after an activity they performed as an educator:

> "I think about something that I used to do in the classroom. When my students would be working in teams, I would walk around with like a green a yellow and a red card. And like if I saw something really great happening in the team, I would give them a green card. And if I maybe heard some struggling or heard something, I might give them like a yellow card or if there was like, either like the team was like being really loud or something like, something was not working in the group, I might give them a red card, and I'd be like, I would just put it on the table and let them sort of talk about what was happening and then I would come back around or maybe eventually I would stop and talk with them about what I'm seeing. So it just reminds me of that."

Rather than a system examining the team as a whole, P9 envisioned a system that could assist teams by way of enabling individual team members to be more informed about their own mental states and contributions:

> "Perhaps prior to the meeting they would set a goal for themselves and basically I'm describing a technology for an individual to manage the efficiency and the productivity of themselves. And those values are kept only for them only for this individual instead of sharing it with the team or team manager."

This approach would theoretically better safeguard user privacy by restricting the availability of their data from others on the team.

Concerns about privacy, equity, and usability of technology to help teams

Despite the potential utility of the new technologies participants suggested, some also voiced concerns about the data that would be collected and how it could be used. If a device measuring the behavior, cognitive states, or collective dynamics of team members existed, these participants correctly noted that this data could be sensitive in nature, and expressed trepidation about such data being collected. For instance, P9 observed,

> "My take on [the question of how a device that could detect how well teams were working together without any limitations] might be switching the role from being a member on the team to being the project manager or the mentor of the project. Because the change of role makes me more want to break into someone's mind and know what they really think of. And I know it this is not ethical. And if I'm the team member, I wouldn't want anyone else to do it."

P10 specifically also raised questions about the equity and usability of future tools and current technology supporting teams more broadly, noting that people who struggle with technology might eventually be left behind if new technology were adopted:

> "We've got to understand there is this gap for people in my age group...I'm not great with technology at all. And I'm seeing people like myself or others who are brilliant leaders, but the technology is holding them back from being able to work with people."

P10 also noted that data from these tools had the potential to counter racial and gender biases if used appropriately, but also had the potential to exclude underrepresented groups as well if they are not adequately considered in the development of a new tool:

> "The sensor to be able to wash your hands. Well, it was based upon white skin. Black skinned people can't use that. You know, just other things that have been built into technology that that is stereotypical or racist. You sit there and you go, "Oh boy, this is going to be really difficult moving forward." Even more so for non-white communities, right, with technology, and just think about ChatGPT and the thinking that has been pulled into it is still predominantly white. And so what are we going to do in order to make it you know, it's just not gonna, not gonna work. We know it. It just, you can't. And that's where it gets really crazy. When technology can be so wonderful. But the limitations are still there. It's only as good as what you put into it."

Summary of Interview Findings

When discussing their experiences working on their respective teams, participants consistently expressed two complementary perspectives on differences between team members, citing incompatible personalities or imbalances in engagement or motivation as challenges to team performance, while highlighting the chance to learn from other team members with different backgrounds or skillsets as one of the most fulfilling parts of collaboration. The need for clear communication and a common understanding across team members also emerged as a crucial element for successful collaboration. Several participants in leadership roles noted that misunderstandings regarding their directions or guidance had led to setbacks in the completion of goals and deliverables, and expressed the desire to mitigate the issues when describing suggestions for future technology or the reasons for their chosen superpowers. Considering how participants used technology to overcome the challenges faced on their teams, most participants engaged in hybrid work, and saw benefits and drawbacks both to working in person and remotely. Remote work allowed teams to meet and have face-to-face conversations easily regardless of geographic distances, but increased reliance on other forms of asynchronous communication technology (such as email or instant messaging applications), which could generate distracting volumes of notifications. In-person interactions, while requiring meeting space and commuting time, provide elements of physical presence without technological overhead, which could increase the depth and effectiveness of communication.

All participants who were interviewed used an array of software applications to streamline their workflows and improve team efficiency. When asked about technologies they wish existed that could help them on their teams, they suggested tools which could detect and respond to team member burnout, allow managers to measure and maintain team morale, and suggest an intervention to resolve conflicts between team members. In addition to such interventions, participants also voiced desires for tools which would allow them to gain insight into team members' mental states, their dynamics with one another, and their relative contributions to the team in order to improve accountability, provide incentive for positive interactions, or allow team leaders to address tension or lack of synergy among the team. Some participants also indicated that a dashboard showing their own mental states and contributions could be useful for setting personal goals and gaining a better understanding of their cognitive and emotional processes during collaboration. Several participants also raised concerns about the privacy, equity, and usability these systems, noting that different power dynamics could exist between team members and leaders and that the usability and usefulness of the system could vary across different groups, especially if there were differences in how the data were used and interpreted.

To explore these further, we incorporated several ideas raised by participants into a set of speculative scenarios illustrating potential implementations of a BCI support system. In the following section, we present participants' reactions to these scenarios, gathered through a "speed-dating" design exercise aimed at uncovering concerns and surfacing design tradeoffs.

7.2.5 Speed-Dating Possible Futures

In order to further explore possible concerns and design tensions arising from the capabilities and use cases participants desired for supportive technology, we presented a series of seven storyboard scenarios depicting possible future implementations of BCI-enabled support tools for teams. These scenarios were presented using a speed-dating design method, which enabled participants to rapidly engage with and react to a range of potential BCI-enabled collaboration tools. Each scenario was crafted to surface possible tensions, tradeoffs, and ethical implications that could arise if such systems were implemented in practice. Participant responses helped assess ways a BCI could help fulfill the needs of teams during collaboration, as well as risks and concerns stakeholders may have about the use of such systems, addressing our second and third research questions.

Setting

Each scenario utilized the same overall setting—namely, a workplace where brain sensing devices have been deployed to assist teams—in order to explore the ramifications of design and implementation decisions, and participants' reactions to these.

The description provided to participants is as follows: "An animation studio is working hard to produce a cartoon series. The production team has tight deadlines, and wants the episodes of the show to be the best they can be before they air. The team works in a smart office of the future, and wear headbands for a support tool that uses brain and behavioral data from team members to measure how well they are working together."

We selected a workplace as our setting because we anticipate this is a context where the need for and impact of a support system for creative collaboration would be greatest, a position corroborated by our discussions with participants. We specifically chose an animation studio because it is a workplace where creative collaboration frequently occurs; where team members can have several different roles and responsibilities, and may be on one or more different sub-teams within the larger organization; which tends to have a hierarchical management structure; and where team members can experience stress and time pressure from deadlines—all important considerations raised by participants during the first experiment session.

Storyboard Scenarios

Using the representative set of 16 "superpowers" from the second experiment session as our basis, we developed seven storyboard scenarios depicting possible design futures for a new BCI system for supporting creative collaboration. Each scenario addressed a particular user need corresponding to a superpower, with several including implementation ideas directly suggested by participants in the first session.

Descriptions of each scenario and the needs/superpowers they address are provided below:

1. **Scenario 1:** Align team effort and output (Figure 7.2). Some animators are putting in **really long hours** and **overtime**, while others **aren't pulling their weight**. The support system indicates this, showing that the **workload** of some team members is **consistently higher** than others'. This awareness enables the team to **align effort** and output and **maintain accountability**.

2. **Scenario 2:** Knowledge of what team members are thinking/feeling (Figure 7.3). Writers and animators are **mocking up storyboards** and a **draft script** for the next episode. Some team members **work together**, while others **work separately**. The team dashboard shows a **map** of the **synergy between team members** based on the brain and behavioral data. It provides notifications when the team is consistently

out of sync and **not working together well**, prompting them to **openly discuss any concerns** before things escalate.

3. **Scenario 3:** Detect and avoid miscommunication/misunderstanding (Figure 7.4). A team member is **giving a presentation** about new directions for the series. Some team members are **confused** by ideas that were raised, but do not **speak up** and ask for clarification. The support tool **detects incomplete understanding** in the moment, enabling **clarification** and **elaboration**.

4. **Scenario 4:** Know how the team can be helped and improved (Figure 7.5). A team dashboard provides **analytics for team dynamics** based on brain and behavioral data. The team uses it to track overall **team dynamics over time**, and how that changes due to **team composition**, **project timelines**, **workload** and **work processes**. It provides **awareness** on how **changes** or **interventions** in their workplace **impact these dynamics**, both within their immediate team and the larger organization.

5. **Scenario 5:** Understand team assets and skills (Figure 7.6). The members of the team perform a **variety of tasks** (mocking up storyboards, drawing video frames, synchronizing frames with audio), each of which **some team members** are **better at** than others. **Team metrics** on the dashboard for some team members are **worse** when they perform some tasks and **better** when they perform others. This makes it easier to know team members' **strengths** and **weaknesses**, and to **allocate responsibilities** and **training more intelligently**.

6. **Scenario 6:** Ability to reduce stress and burnout (Figure 7.7). After **working hard** and putting in **overtime** to meet **upcoming deadlines**, the dashboard indicates that multiple team members are **getting burnt out**. It recommends they take a break, and suggests **quick and fun activities** they can do **together** to **take their minds off work**.

7. **Scenario 7:** Build and maintain trust/camaraderie among the team (Figure 7.8). Marcy **isn't a big fan** of working with Jake. He's **very extroverted** and **always strikes up a conversation** with everyone on the team, but is also one of the team's **highest performers**. Marcy is **more introverted**, and **struggles to raise her ideas** or **feel good about her work** when Jake is in the group. However, she notices on the team dashboard **that there are times when Jake has low focus** and **seems to struggle too**, which makes her **feel a bit better** about him and **enables her to work more confidently** with him.

Analysis and Results

For each scenario, participants were asked to answer the following questions, and to voice their thoughts aloud as much as possible:

1. Consider this storyboard showcasing a new technology for teams. Do you think the solution presented meets the needs of the team in this scenario? Why or why not?

2. What if the system provides a real-time indicator/notification vs. allowing after-the-fact reflection?

3. Consider the visibility of the data the tool collects. Does anything change if the data is private vs. shared (with team, with manager)?

4. Would your perspective change if you were the team leader/manager vs. being a team member? Would your answers to previous questions change?

5. What potential concerns, if any, do you think might arise in this scenario?

Additionally, we also asked the following question about diversity and equity during the first scenario, and noted that participants might consider this in future scenarios:

6. What if the team member is underrepresented in the group? Does anything change?

The transcripts for the speed-dating sessions were analyzed via inductive open coding to uncover any common themes. Results for each of the scenarios are summarized below.

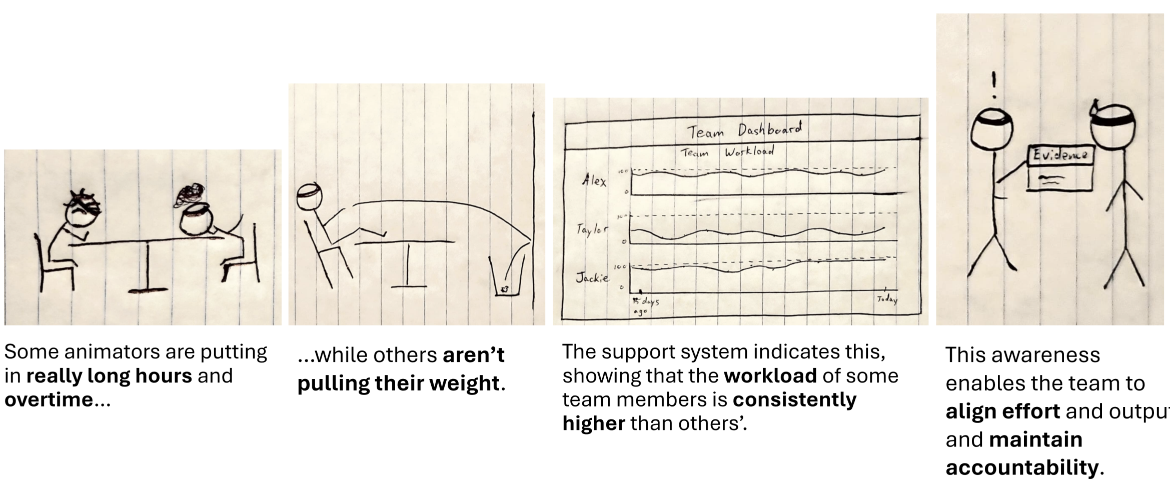
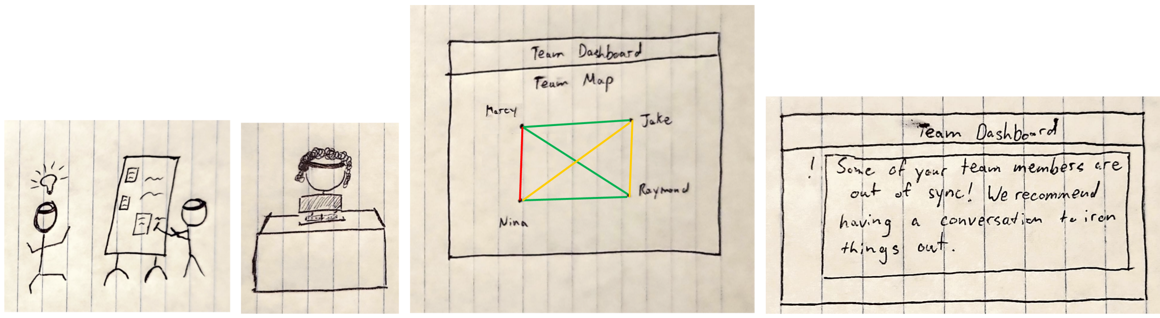


Figure 7.2: Scenario 1: Align team effort and output

Scenario 1: Align team effort and output Participants were about evenly polarized in their perspectives regarding this scenario, with half of the participants believing the solution presented met the needs of the team, while most others did not (one participant believed that the solution met some needs of the team, but not others). Participants were similarly split between whether the data from the system was best presented in real time versus after the fact—with a slight preference for real-time updates—as well as perspectives on the visibility of the data, with slightly more than half in favor of keeping the data restricted between individuals and managers, and the others mainly favoring visibility to the whole team (either anonymized or not). Finally, participants were similarly divided on whether their perspective on the system would change if they were managing a team versus a team member, with slightly more than half stating it would (team members might prefer that the data be private and not shared with other team members, but managers could use it as a tool to help determine if an employee is slacking, or as a basis for checking in about their well-being if workload is consistently high).

Participants raised concerns that the data could be used to enforce a particular model of productivity or teamwork, and that workload might not be enough for a full picture of team member performance—some people could have low workload but high output, and some might be the opposite. Additionally, the system also did not appear to factor in the different roles and responsibilities of team members; it is possible that team members in some roles would be expected to exhibit higher or lower workload than members in different roles. The visibility of the data could also lead to unhealthy competition or reduced morale among team members, if team members try to outperform one another on the basis of workload metrics. Some team members could also try to game the system, attempting to make sure their workload is consistently high, or lower it to meet the levels of others. Participants were also concerned that the workload values observed could be is misinterpreted or misattributed; it is not clear that the values observed are due to the nature of the work as opposed to external factors. Thus using the data from the tool as the sole basis for managerial decisions is likely unwise. Finally, participants expressed concerns about the privacy of the data, especially if the data were shared explicitly with others on the team.

Scenario 2: Knowledge of what team members are thinking/feeling Participants were similarly divided in their responses for the second scenario. Slightly more than half of the participants believed the solution presented met the needs of the team, and were likewise divided regarding the granularity of the data, with half preferring updates in real time and three preferring after-the-fact reflection, and one participant seeing merits to both approaches. Participants were split on preferences for the visibility of the data, with equal portions in favor of keeping the data restricted between individuals and managers, visible only to the manager, and visible to the whole team (either anonymized or not). Finally, participants were similarly divided on whether their perspective on the system would change if they were managing a team versus a team member, with half stating it would; managers could use the data to help team members navigate potential interpersonal friction, while



Writers and animators are **mocking up storyboards** and a **draft script** for the next episode.
Some team members **work together...**

...while others **work separately.**

The team dashboard shows a **map** of the **synergy between team members** based on the brain and behavioral data.

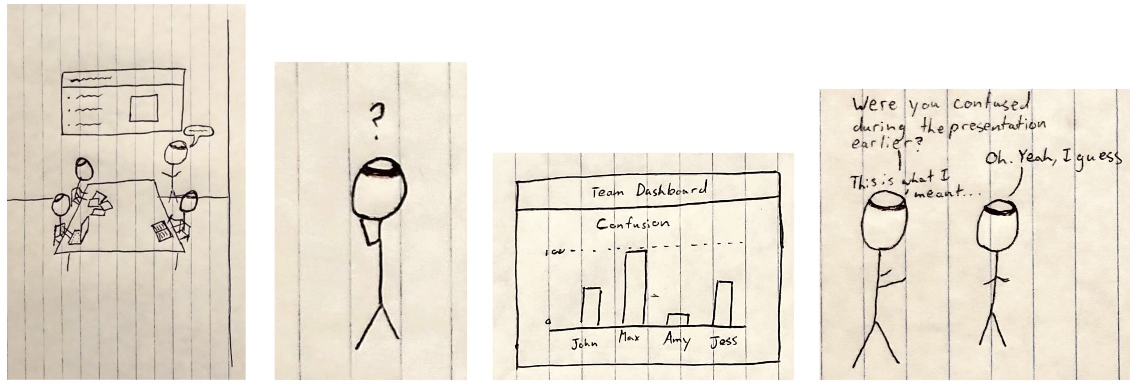
It provides notifications when the team is consistently **out of sync** and **not working together well**, prompting them to **openly discuss any concerns** before things escalate.

Figure 7.3: Scenario 2: Knowledge of what team members are thinking/feeling

team members might use it to clarify their relationships with others with whom they have poor synergy, or prioritize interactions with those where the system indicates synergy is optimal.

Regarding potential concerns, one participant worried that the system could enable micromanaging by the team manager if they were constantly monitoring the synergy of team members. There were also concerns that the data could be used to enforce a particular model of productivity or teamwork, because some people might not be as comfortable in social situations, and might work better alone for certain tasks while periodically updating the rest of the team on progress. The data does not give a full picture of team interactions or performance; it might be difficult to understand *why* the synergy between some people is low, especially if their subjective perceptions of their interactions are positive. Participants were also concerned that the data could be misinterpreted or misattributed, or that the feedback from the system could be inaccurate, since it is unclear how "synergy" was being defined and measured. It is also unclear what impact a team member working separately from the others would have on the the synergy levels depicted by the system. Participants also expressed concerns about the privacy of the data, and that showing low synergy between some people and not others could lead to negative feelings between team members.

Scenario 3: Detect and avoid miscommunication/misunderstanding There was much more consensus in participant perspectives on the solution presented in the third scenario, with all participants agreeing that it met the needs of the team. All but one participant was in favor of the system working in real time to capture team member confusion, with nearly half additionally supporting after-the-fact reflection (so that the presenter or manager could see what data from the audience looked like after the presentation). Nearly



A team member is **giving a presentation** about new directions for the series.

Some team members are **confused** by ideas that were raised, but do not **speak up** and ask for clarification.

The support tool **detects incomplete understanding** in the moment...

...enabling **clarification** and **elaboration**.

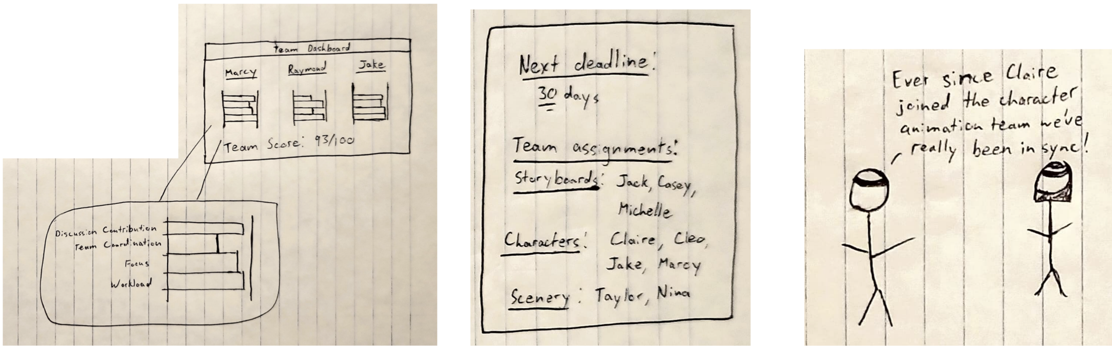
Figure 7.4: Scenario 3: Detect and avoid miscommunication/misunderstanding

all participants believed that the data should be anonymized, but visible to the presenter, and a slightly majority stating their answers would be unchanged regardless of if they were a team member or leader.

Participants raised concerns mostly about the social stigma surrounding the intervention enacted in the scenario, and whether a team member exhibiting hight confusion would lead to an uncomfortable confrontation with the presenter after the presentation to resolve it. The presenter responding to a team member's confusion in public with other members of the audience present could lead them to feel dumb or called out. Finally, there were also some concerns that the results could be inaccurate, or due to some factors unrelated to confusion, such as lack of sleep.

Scenario 4: Know how the team can be helped and improved Participants unanimously believed that the solution presented in scenario 4 met the needs of the team, and most were in favor of a system that allowed for after-the-fact reflection (i.e., one that presented a daily or weekly average of each metric) versus one that updated frequently in real time. There was greater diversity in preferences for the visibility of the data; half of participants felt that the data should be kept between managers and individuals, while others were in favor of the data being shared with the whole team, strictly visible to either individuals or managers, or the team as a whole. Participants tended to feel that their perspectives would be different if they were a manager instead of a team member in this scenario, noting that team members might be able to use the data to advocate for reassignment to a different role, or demonstrate to leadership that a particular team composition is not working well.

Half of participants had no concerns for this scenario. Some concerns raised by the



A team dashboard provides **analytics for team dynamics** based on brain and behavioral data.

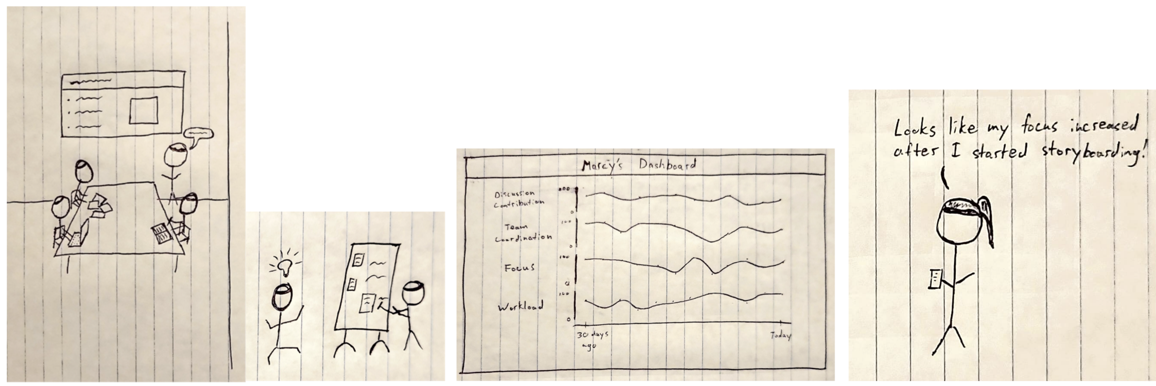
The team uses it to track overall **team dynamics over time,** and how that changes due to **team composition, project timelines, workload** and **work processes**.

It provides **awareness** on how **changes** or **interventions** in their workplace **impact these dynamics**, both within their immediate team and the larger organization.

Figure 7.5: Scenario 4: Know how the team can be helped and improved

remainder were that the metrics measured by the system might not necessarily equate to performance on the team, and might not accommodate different learning or collaboration styles. It is also difficult to know whether the observed dynamics are due to factors related to work or external to it, or what the relationship between the dynamics measured by the system and productive output may be; it is possible people could still be productive even if their data is different from others', and thus using the data from the tool as the sole basis for managerial decisions would also be concerning if it occurred. Participants were also concerned that team member morale could be reduced if the brain metrics of the team become "worse" if someone is added, or if members see that others' brain metrics are "better." Privacy was also a concern.

Scenario 5: Understand team assets and skills There was a similar degree of consensus in participant responses for the fifth scenario. All participants thought that the solution presented met the needs of the team, and were largely split between a system able to provide after-the-fact metrics versus one able to work both after the fact and in real time. Participants were likewise evenly split in their preferences for the visibility of the data provided by the system, with half of participants preferring that data only be shared between individual team members and a team leader, and half that it be shared with the whole team—a majority of whom preferred it not be anonymized for other team members, so that teams could make adjustments based on others' capabilities and be mindful of the fact that all team members have different strengths and weaknesses. Most stated they would have a different perspective depending on whether they were a team manager or leader versus a team member, with managers ideally leveraging the data from the system to ensure employees are assigned tasks suited to their individual strengths, and employees using the data to present a



The members of the team perform a **variety of tasks** (mocking up storyboards, drawing video frames, synchronizing frames with audio), each of which **some team members** are **better at** than others.

Team metrics on the dashboard for some team members are **worse** when they perform some tasks and **better** when they perform others.

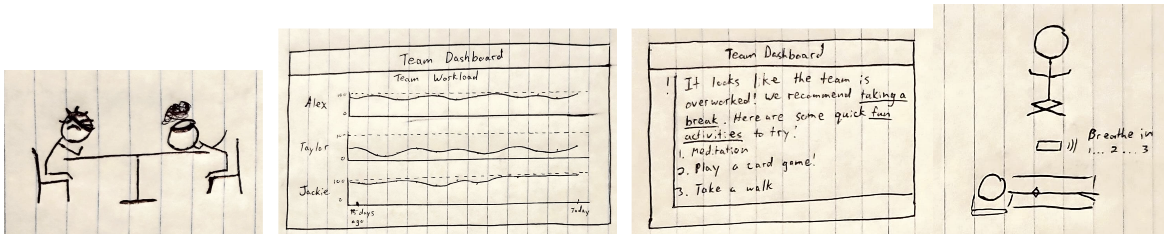
This makes it easier to know team members' **strengths** and **weaknesses**, and to **allocate responsibilities** and **training more intelligently**.

Figure 7.6: Scenario 5: Understand team assets and skills

case to managers for taking on different responsibilities if they are struggling in their current role.

Participants were concerned that if the team manager could overemphasize the data measured by the system when assessing team member performance—the data is important, but it should not be the only thing used to evaluate team effectiveness; outcomes and deliverables are also important. There were also concerns about hurt feelings if people were unhappy with their brain data (e.g., if they changed teams because their metrics were low and the metrics didn't change in response), or that team members could become discouraged or compare themselves unfavorably if the data were shared with others. One participant also noted that making decisions about team assignments based on the data could lead to unfavorable outcomes for team members, since they might be good at a task or role according to the metrics the system measures even if they are not passionate about it, or vice versa. Data privacy was also a concern.

Scenario 6: Ability to reduce stress and burnout Participants were largely in agreement about their perceptions of the solution presented in the seventh scenario, with nearly all indicating that it met the needs of the team. They largely saw merits in implementations of the system able to provide updates in real time as well as periodic reports allowing for after-the-fact reflection, and generally preferred that the data collected only be visible to managers, not team members. A majority of participants also stated that their perspectives on the system would differ depending on whether they were team members or managers, with managers able to utilize the data to reflect on how they are managing their employees, and ensure that they are providing enough support to mitigate the risk of employee burnout.



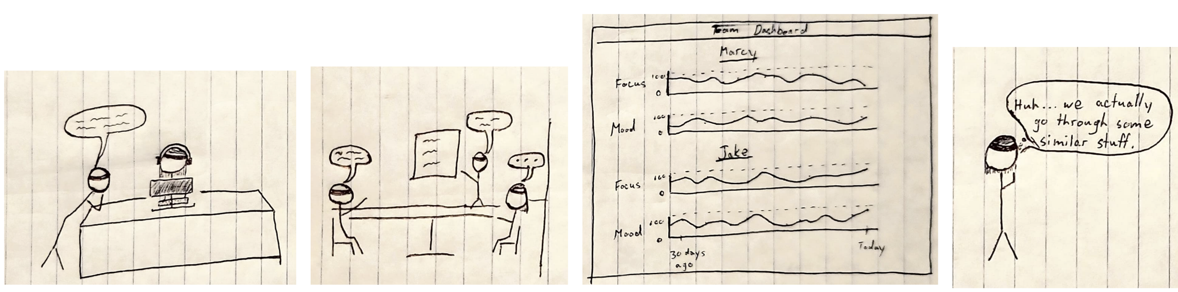
After **working hard** and putting in **overtime** to meet **upcoming deadlines**...

...the dashboard indicates that multiple team members are **getting burnt out**.

It recommends they take a break, and suggests **quick and fun activities** they can do **together** to **take their minds off work.**

Figure 7.7: Scenario 6: Ability to reduce stress and burnout

Participant concerns for this scenario centered around the reliability of the measurement and the feasibility of the data. One participant noted that team members could start at their positions burnt out (and might therefore have an incorrect baseline), or the system could be measuring burnout due to factors outside of work, and might not differentiate between different causes of burnout. Other participants were concerned that some team members could try to game the system to make it recommend breaks more frequently. If breaks are collective, recommending a break could be distracting for team members who are focused and on-task. Other participants noted that depending on the organization and work environment, breaks sometimes are not feasible, especially if the team is operating under a tight deadline. Participants were also concerned about the privacy of the data, and whether the intervention should change if the same person was consistently burnt out.



Marcy **isn't a big fan** of working with Jake. He's **very extroverted** and **always strikes up a conversation** with everyone on the team, but is also one of the team's **highest performers.**

Marcy is **more introverted**, and **struggles to raise her ideas** or **feel good about her work** when Jake is in the group.

However, she notices on the team dashboard **that there are times when Jake has low focus** and **seems to struggle too**...

...which makes her **feel a bit better** about him and **enables her to work more confidently** with him.

Figure 7.8: Scenario 7: Build and maintain trust/camaraderie among the team

Scenario 7: Build and maintain trust/camaraderie among the team Participants near unanimously agreed that the solution presented met the needs of the team in the seventh scenario, but were near evenly divided on how granular they thought the data should be, with similar degrees of preference for whether the data should be shown in real time, aggregated after the fact, or both Participants likewise disagreed about how visible the data should be, with near equal degrees of preference for the data being shared with individuals and managers, managers only, a team member's immediate work group, and the whole team (either anonymized or not). Participants likewise agreed that their perspectives would change if they were team managers, noting that while team members might prefer to either keep the data completely private, or else view others' data while keeping their own protected, managers might take on the role of intermediaries, noticing in the data if team members are struggling and helping them positively reframe their relationships with coworkers.

The chief concern participants raised about this scenario was hurt feelings or reduced morale among team members. Most noted that while the outcome was ultimately positive in this case, having everyone's data visible could encourage team members to look for shortcomings in others. Furthermore, giving managers access to employee data could encourage or enable micromanaging; if the data were visible to team members, they might ascribe qualities or mental states to coworkers that are incorrect (e.g., that someone is dumb or depressed when they are not). The privacy of the data was also a concern.

What if the team member is underrepresented in the group? Just over half of participants indicated that their thinking about the system would change if the team member featured in the scenario was part of an underrepresented group, with those who said it would not specifying that the metrics used by the system should not be the primary basis for managerial decisions, such as firing, hiring, or providing raises. Participants who stated that the team member's underrepresented status was an important consideration raised concerns that the data output by the system could be used to reinforce existing stereotypes or contribute to othering, or that the data itself could exhibit biases based on gender, race, age, or other demographics.

7.3 Discussion and Future Work

The results of our study provide insights into the potential role of BCIs in facilitating effective teamwork, highlighting both opportunities and challenges. Here we discuss answers to our research questions, implications of the results for the design of a BCI system to support collaboration, and future work exploring the possible implementation of such a system.

7.3.1 RQ1: What unmet needs do stakeholders have during collaboration?

Across our study, participants consistently voiced a desire for tools that could enhance shared understanding, foster communication, and mitigate friction caused by differences in work style or motivation. While technology has already begun to support remote collaboration and task management, participants expressed that these tools do little to reveal how their teams are functioning on a cognitive and emotional level. In interviews with participants, miscommunication and misalignment of expectations were frequently cited as barriers to effective collaboration, especially in circumstances where team members refrained from openly expressing confusion, or were otherwise unaware miscommunication had occurred. This is supported by the fact that "know how the team can be helped and improved," "detect and avoid miscommunication," and "align team effort and output" were among the most prioritized superpowers during card sorting, reflecting an unmet need for greater shared understanding of team dynamics, especially between team members and team leaders.

Cognitive overload and stress also emerged as concerns, with many participants noting that burnout negatively impacted both individual and team performance. Participants expressed a desire for insight into the well-being and workload of their team members, not only to prevent burnout, but to ensure more equitable distribution of responsibilities, as some team members reported taking on disproportionate amounts of responsibility. However, these needs were often tempered by concerns about privacy and the ethics of surveillance, highlighting a need for systems that can support these goals without overstepping personal boundaries or compromising trust.

7.3.2 RQ2: In what ways could brain-computer interfaces help fulfill these needs?

BCIs have the potential to provide real-time insights into cognitive and emotional states, enabling more adaptive and responsive collaboration. Indeed, as described above, participants voiced several challenges and unmet needs over the course of their experiment sessions, such as detecting stress, imbalances in workload, or differences in shared understanding, which may be measured by such systems in the future. When asked to imagine how a future technology which could detect how well teams were working together might improve their team interactions, participants envisioned a variety of use cases: dashboards showing levels of engagement during meetings, alerts signaling when teammates are confused, or visualizations that help teams understand their social dynamics over time.

Such ideas could feasibly be implemented as BCI support systems for teams (and are scenarios we highlighted via storyboards in the second session, incorporating these ideas from participants): a system might detect when individuals were disengaged or experiencing cognitive overload, thus allowing for timely interventions. A system capable of non-intrusively monitoring engagement levels could help team leaders adjust communication strategies to ensure discussions remain productive and inclusive. Additionally, BCIs could assist with workload distribution by identifying early signs of burnout, prompting proactive adjustments to task assignments. By synthesizing neurophysiological data at the team level, a well-designed BCI could enhance awareness of group interactions, improving coordination and decision-making without disrupting workflow. Crucially, while engaging with the storyboard scenarios depicting such implementations, participants saw merit both for systems capable of providing feedback or alerts in real time as well as systems which provide summaries reports for future reflection or consideration, both use cases that BCI systems could accommodate.

7.3.3 RQ3: What concerns do stakeholders have about BCIs being used to support teams, and what risks do they foresee?

Despite excitement about the possibilities of BCIs, participants expressed serious concerns about how such technology could be implemented in practice. Chief among these related to privacy and autonomy: participants worried about how their brain data might be used by managers or institutions, particularly if it were linked to performance evaluations or used to make decisions about promotion, hiring, or firing. There was also skepticism regarding the accuracy of neurophysiological signals as indicators of engagement, given the complexity of cognitive states and the potential for misinterpretation. There was a strong consensus that access to brain data should remain under the control of the individual, with transparency about what is collected, how it is used, and who can see it.

Equity and inclusivity also surfaced as important issues. Participants expressed concern that BCIs, if not designed with diverse populations in mind, could exacerbate existing biases or exclude certain groups (e.g., older individuals or those from underrepresented backgrounds in tech). Others noted that such tools could become burdensome or inaccessible for individuals less comfortable with new technology, raising usability and adoption concerns.

Finally, some participants worried that such a system could inadvertently introduce new forms of surveillance, reinforcing managerial oversight in ways that might pressure employees to conform to predefined behavioral expectations. Participants imagined futures where brain data could be used coercively or punitively, and expressed discomfort with systems that inferred intention or cognition without their explicit consent.

7.3.4 RQ4: How could such risks and adverse impacts be minimized over the course of design and development?

To responsibly design BCIs for teamwork, it is imperative to adopt a **user-centered** and **participatory** design approach from the earliest stages. Based on our findings, we developed guidelines for the ethical design and deployment of these systems, which we further outline in Section 8.3.

To summarize, we assert that use of the system should be fully voluntary, with users retaining control over whether and how they engage with the system, including which data (if any) are shared and with whom. Consent must be dynamic and revisable, allowing individuals to opt out or modify their settings at any time without facing repercussions or stigma. Data practices should adhere to a principle of minimalism: raw brain signals should never be stored, and only the smallest necessary set of derived metrics should be computed. These metrics must be purpose-bound, anonymized, and retained only for a limited, clearly defined duration aligned with user goals and project needs.

Equally important is a commitment to transparency and shared understanding among all stakeholders. Users, team leads, and administrators must be clearly informed about what the system captures, how it works, and what the data means—ideally through collaborative onboarding or participatory design practices. Feedback should be framed to support well-being and team functioning rather than serve as a tool for evaluation or compliance. To avoid coercion, the system should be useful even when adopted by only a subset of team members. Finally, neurophysiological data must never be used in isolation to guide managerial decision-making; instead, it should be interpreted alongside behavioral observations, team discussions, and broader organizational context to ensure fair, holistic, and informed action.

By adhering to these principles in the design process, developers can build tools that are not only powerful, but equitable and respectful of the individuals they aim to support.

7.3.5 Future Work

This study lays the groundwork for the ethical development of BCI systems to support collaborative work, but many avenues remain for future exploration. First, a broader sample of participants from different industries and cultures could yield additional insights and help test the generalizability of these findings. Second, the design and evaluation of low-fidelity prototypes based on participant feedback could be used to further validate acceptability and refine technical requirements before implementing real-time systems.

Additionally, studies could investigate the *longitudinal* use of such systems: how do team members' perceptions change over time? Does regular feedback improve team cohesion, or introduce fatigue or resistance? There is also room to explore hybrid approaches that blend BCI input with behavioral data or self-reports to triangulate team dynamics while offering users greater interpretive context and control.

Ultimately, the development of BCIs for collaboration will require the convergent efforts of neuroscientists, HCI researchers, ethicists, designers, and end users. Integrating these unique perspectives throughout the design process will ensure that these systems meaningfully enhance teamwork while safeguarding user autonomy and well-being.

7.4 Conclusion

This study highlights both the opportunities and challenges of implementing brain-computer interfaces to support collaborative teamwork. Through engaging team members and leaders in user-centered design interviews and brainstorming sessions, we identified key unmet needs, such as enhancing shared understanding, preventing miscommunication, and managing cognitive overload and stress. Participants envisioned BCIs addressing these needs by providing insights into cognitive states and team dynamics, enabling proactive, supportive interventions. However, significant concerns about privacy, autonomy, equity, and usability also emerged, underscoring the importance of ethical guidelines and careful design. By proposing concrete ethical guidelines for voluntary participation, minimal and transparent data collection, supportive feedback mechanisms, and contextual decision-making, this research lays a foundation for developing BCIs that effectively enhance teamwork while prioritizing the well-being and autonomy of individual users.

Chapter 8

Ethics and Broader Impacts

The emergence of brain-computer interfaces within collaborative and workplace contexts introduces pressing ethical questions that demand proactive engagement. While BCIs have historically been deployed in clinical and accessibility-focused applications, more recent work has expanded to focus on monitoring and augmenting cognitive and affective processes in a wider variety of contexts, including in teams and organizations. My dissertation investigates both technical and user-centered dimensions of this shift, particularly with regard to making BCI systems more accessible and equitable, and supporting team dynamics in future workplaces. This chapter focuses on ethical considerations and broader societal impacts, drawing on prior work in neuroethics and empirical data from my own user-centered study on BCIs for teamwork.

As the use of BCIs extends beyond therapeutic interventions and into the realm of performance augmentation, questions surrounding autonomy, privacy, consent, and equity become increasingly salient. Scholars and international frameworks such as the IEEE Neuroethics Framework [147] and IEEE Ethically Aligned Design [148] have identified core principles to guide ethical BCI development, including mental privacy, transparency, and fairness. However, while some empirical studies have begun to examine stakeholder perspectives, particularly in general workplace contexts, grounded empirical insights into how stakeholders perceive and prioritize ethical concerns in team-based or collaborative BCI settings remain limited. My research contributes to this discourse by offering a detailed, stakeholder-informed perspective on the ethics of collaborative BCI systems.

This chapter is organized as follows: First, I survey key themes in the neuroethics literature and ethical BCI design, highlighting relevant values and theoretical orientations. I then explain how my own research addresses gaps in this body of work, showing how participants imagined, critiqued, and contextualized BCIs within collaborative settings. Then, I integrate these insights from my work and the literature to provide guidelines for the design of ethically responsible BCI systems for collaboration. Finally, I examine the broader impacts of this work, focusing on how it contributes to making brain signal research more accessible and reproducible within the HCI community, and how ethically designed BCIs can support collaboration.

8.1 Prior Work in BCI Ethics and Neuroethics

Neuroethical discussions of BCIs frequently center on core values such as autonomy, mental privacy, responsibility, and identity. These values not only reflect long-standing ethical principles but are also inflected in new ways by the particular affordances and risks of brain-computer interaction technologies. Several scholars have organized their analyses around these value domains, making it possible to synthesize insights across different traditions and disciplines.

Autonomy is one of the most frequently discussed concerns in BCI ethics. The potential for BCIs to influence, predict, or bypass user intention raises fundamental questions about self-determination and agency. Passive BCIs, for example, can infer affective states or cognitive load without requiring the user's active engagement, raising questions about whether users have meaningful control over how they are represented or interpreted by the system. Emerging work on multi-person brain-to-brain interfaces [128], where brain signals can be transmitted and received by multiple users, raises further questions about the preservation of individual intentionality and free will. In this context, minds operate collectively as part of a network, and individual users may not be able to easily discern the precise origins of the data they receive, or might transmit data unintentionally, blurring the boundaries of consent if user output depends in part on the inputs from others. In the workplace, autonomy may also be compromised by unequal power dynamics—for instance, when organizational pressure implicitly compels workers to adopt neurotechnologies [40, 203].

However, the literature also suggests that BCIs can enhance autonomy in some contexts. For example, when designed with adjustable transparency and user-configurable settings, BCIs may empower individuals to reflect more deeply on their cognitive and emotional states [96]. In group settings, shared neurofeedback has the potential to help team members identify miscommunication or interpersonal tensions early [109, 236], enabling more intentional collaboration. Thus autonomy is not just a matter of individual choice but is shaped by context, and highlights the importance of designing systems that allow for negotiated, ongoing consent and support user self-determination [109].

The issue of mental privacy builds upon concerns about autonomy but introduces its own set of challenges. Unlike traditional biometric or behavioral data, neurodata can be deeply intimate, potentially revealing aspects of a user's intentions, emotional state, or cognitive processes, raising concerns about who has access to this information and how it might be used. Some scholars advocate for the recognition of mental privacy as a distinct human right, given the unprecedented access BCIs provide to internal states [31, 375]. Workplace applications intensify this concern, especially when users are not fully aware of the kinds of inferences being made from their brain activity or when data is accessible to employers [219, 236].

Responsibility emerges as a complex issue, particularly in systems where BCIs augment or mediate action. If a decision is influenced by neuroadaptive interfaces, assigning accountability becomes more difficult. Questions arise around whether users are fully responsible for decisions that were facilitated or filtered through BCIs—if such devices are only able to measure a small portion of the user's full brain state, can the user ever be held liable for downstream consequences? [203] Moreover, system designers and organizations share ethical responsibility for creating environments where these tools are used. Ethical frameworks that emphasize relational and shared responsibility, rather than individual blame, are particularly relevant [317, 341].

Another important dimension is identity and selfhood. Using neural activity for feedback or performance evaluation can influence how users see themselves, their roles, or their emotional and cognitive capacities. Such feedback might reinforce certain narratives—such as being a "focused" or "high-performing" teammate—but could also produce tension or alienation, especially if users feel misrepresented by the system [109]. These concerns are often framed in terms of narrative identity: the capacity to construct and sustain coherent self-understandings through time [271]. Still, identity-oriented frameworks also underscore the positive potential of BCIs to support user reflection. Systems that allow users to explore fluctuations in attention or emotional state may foster greater self-awareness and personal growth [264, 271], provided they avoid essentializing or determinist interpretations of neural data.

Finally, equity and inclusivity also emerge as key ethical concerns, particularly regarding access to BCI technologies and differential impacts across social groups. Studies emphasize the risks of marginalization for users with less technological fluency or for those whose neurodata does not align with normative expectations. There is also concern that BCIs might entrench workplace hierarchies or be differentially imposed across roles or sectors [203, 326].

Across these works, ethical frameworks vary: some adopt a rights-based approach [31], others draw on care ethics [341], relational autonomy [330], or empirical ethics [40]. Despite these differences, a shared concern emerges: BCIs pose novel challenges to dignity, equity, and control that cannot be addressed by technical design alone. A recurring critique is that ethical discussion must move beyond high-level speculation to address the everyday realities of BCI use. Scholars call for empirical, participatory, and interdisciplinary approaches that incorporate stakeholder values and lived experience [40, 60, 264].

Taken together, these concerns suggest that ethical BCI development must attend to both the technological capacities of systems and the relational, situated contexts in which they are used. The following section builds on these insights by examining how participants in my research articulated their own values and ethical concerns about collaborative BCIs.

8.2 Ethics in Context: Stakeholder Perspectives on the Development of BCIs for Workplace Teams

This dissertation contributes to ongoing neuroethical conversations by grounding ethical principles—such as autonomy, privacy, and responsibility—in the lived perspectives of potential end-users. Participants in my user-centered study examining stakeholder perspectives on BCIs for collaborative contexts (Chapter 7) surfaced both expected and underexplored ethical dimensions of collaborative BCI use. Their responses affirm many of the concerns raised in the prior literature, but also extend these lines of ethical inquiry to the affective and interpersonal dynamics within teams.

Participants voiced clear concerns about autonomy and coercion, echoing existing discussions. Several participants noted that even if it were aggregated or anonymized, sharing data depicting the cognitive and affective states of team members could introduce subtle pressures to conform. For instance, some participants expressed worry that making mental workload levels of all team members visible could lead to a kind of peer pressure, which would encourage them to either attempt to maximize their workload levels to meet expectations set by supervisors, or moderate their workload to match the rest of the team. This highlights how consent is not only an individual decision, but embedded in social relationships and power structures. Others emphasized the importance of being able to configure what data is shared and with whom, stressing the value of contextual autonomy and selective self-disclosure.

Privacy and mental transparency were also key themes. Participants consistently raised the fear of involuntary exposure, particularly in contexts where emotional data might be shared in real time. For example, several participants each described scenarios in which a teammate might appear confused or disengaged on a shared dashboard, without any room for that data to be qualified or contextualized. Such moments were seen as potentially embarrassing or damaging to morale. At the same time, others expressed excitement about being able to "see how the team is doing" in a more immediate and nuanced way, suggesting that with careful design, neurodata could foster empathy and group awareness.

Responsibility and accountability were addressed implicitly in participant reflections on misinterpretation. Concerns were raised about team leaders using neurodata to draw premature or incorrect conclusions about individual contribution or engagement. For example, participants noted that emotional or attentional states fluctuate for reasons unrelated to team performance, and that BCIs could easily amplify bias if their outputs were over-trusted. These reflections highlight the need for shared responsibility in BCI-mediated teamwork, where system developers, users, and organizational leaders collaborate to establish clear norms for how brain data should be interpreted and used to inform action.

Despite these ethical concerns, participants saw promise in BCIs as tools for improving communication and reducing misalignment. Several mentioned that teams often struggle with differences in member understanding or engagement, or latent frustration that goes unspoken until it becomes disruptive. A BCI-based system that gently indicates these sources of divergence might allow teams to address issues earlier, when they are more manageable. Others noted that BCIs could help individuals monitor their own cognitive states and set personal goals for focus or workload regulation, which could also support narrative identity by enabling users to better understand and author their mental experiences. Participants also speculated that objective neural data might offer a counterbalance to racial or gender biases in workplace assessments, helping to validate individuals' contributions or cognitive engagement when such factors might otherwise be overlooked or undervalued.

8.3 Ethical Guidelines for the Responsible Use of BCIs for Collaboration

These findings from my work suggest that ethical design of BCIs for collaborative contexts must consider not only the dyadic relationship between the user and system or employer, but the broader network of social dynamics and interpretations that shape how neurotechnologies are experienced and understood within teams.

Drawn largely from a care ethics perspective, which prioritizes healthy interpersonal relationships and well-being, and integrating the insights outlined in prior sections, I propose the following guidelines for the design of ethically responsible BCI systems for collaboration. These consist of six interrelated principles:

1. **Agency and Configurable Consent** — Team members should retain full agency over whether and how they engage with a BCI system. They must be able to choose whether to use the system at all, which data (if any) is shared, and how visible it is to others. Consent should be ongoing, not a one-time agreement, and users must have the ability to opt out or adjust their settings at any time without penalty or stigma.

2. **Minimal and Purpose-Bound Data Collection** — Organizations must err on the side of collecting and retaining as little data as possible. Raw brain data should never be stored, and only the minimal set of derived metrics needed to support intended functionality should be generated. These should be anonymized and aggregated by default, and stored only for a limited, clearly defined duration, such as the length of a project or a user-specified timeframe.

3. **Transparency and Stakeholder Comprehension** — All stakeholders—including employees, team leads, and system administrators—must be clearly informed about what data will be collected, what metrics are calculated, what those metrics mean, and how they will be used. The benefits and risks of using the system must be openly discussed, ideally as part of a collaborative onboarding or co-design process.

4. **Support-Oriented Feedback Design** — BCIs should function as tools that enhance collaboration and team well-being, not as mechanisms for performance monitoring or behavioral enforcement. Feedback should be constructive, team-centered, and framed to support mutual understanding and the overall health of the team. Systems that foster interpersonal trust and shared responsibility are more likely to be ethically sustainable.

5. **Optional and Non-Uniform Participation** — The system's usefulness must not depend on universal adoption within the team. Requiring full participation compromises agency and may coerce individuals into opting in to maintain alignment with teammates or management. Instead, systems should be designed so that partial usage still offers value.

6. **Contextual and Collaborative Decision-Making** — Neurodata should never serve as the sole basis for managerial decision-making, especially in punitive or evaluative contexts. Instead, it should be used as one form of context among others, triangulated with behavioral observations, team dialogue, and other organizational knowledge.

Together, these principles aim to create conditions where BCIs can serve as empowering, reflective tools that improve team cohesion and communication without introducing new burdens or inequalities.

8.4 A Motivating Example

To illustrate how these guidelines might be enacted, consider an optimal implementation of a team-facing BCI system in a workplace setting:

Imagine a cross-functional project team in a mid-sized technology firm. Team members opt into using a lightweight BCI headband during project sprints. Each user configures their settings individually, choosing whether to share their data with others, and if so, which categories: emotional valence, cognitive workload, and/or focus. Shared data appears on a central dashboard only in aggregate form (e.g., "team stress trending upward"), color-coded for clarity. Raw neural signals are never stored, and all derived metrics are deleted at the end of the project.

The system was co-developed with employee input, and onboarding included a values-centered workshop where stakeholders discussed how the tool should be used—and not used. Feedback appears periodically and unobtrusively in team retrospectives, highlighting patterns rather than individual anomalies. A rising team stress score might trigger a shared reflection or lead to minor scheduling adjustments, rather than disciplinary action. Individual users can access their own historical data if they choose to reflect on personal patterns over time.

Importantly, the system continues to function even if only a subset of team members use it. This allows for staggered adoption and respects differing comfort levels. Throughout deployment, an internal review board evaluates outcomes and updates system policies in response to user feedback. Managers are explicitly prohibited from using the data for individual performance evaluation; rather, it serves to enrich ongoing conversations about workload, communication norms, and team well-being.

Such a system does not eliminate ethical risks, but does reflect a careful balancing of power, agency, and insight. It operationalizes ethical commitments not through values alone, but through specific, embedded design features and workplace practices that prioritize respect, care, and collaboration, as well as exemplifies under what conditions using BCIs in team contexts is justifiable.

8.5 Broader Impacts

This dissertation contributes to a range of broader impacts across domains including HCI research reproducibility, the accessibility of brain data for design, ethical technology development, and public understanding and engagement.

8.5.1 Advancing Research Reproducibility and Accessibility in HCI

One of the most immediate impacts of this work lies in improving the reproducibility and accessibility of brain-computer interface research in the HCI community. As noted in our work developing an experiment model for conducting reproducible BCI research (Chapter 3, [276]), reproducibility in brain signal research within HCI has been hindered by inconsistent reporting practices, methodological diversity, and a lack of unified documentation standards. The development of structured experiment models and recommendations for transparent reporting directly addresses these concerns. By adopting reproducibility-friendly approaches and clearly reporting data acquisition, analysis pipelines, and experiment design, this dissertation contributes to a more robust and cumulative body of research.

Moreover, the introduction of BrainEx (Chapter 4, [139]) lowers the barrier to entry for exploring brain signal data by enabling intuitive visual exploration and search functionality. By facilitating use by non-experts and emphasizing usability, tools like BrainEx open up possibilities for broader participation in neurotechnology research and design. These contributions support the National Science Foundation's societally relevant outcomes of advancing discovery and innovation through shared infrastructure and cross-disciplinary collaboration, as well as fostering more inclusive research ecosystems [250].

8.5.2 Societal and Workplace Impacts of BCIs for Teams

From a societal perspective, the potential deployment of BCIs in team settings raises both significant opportunities and concerns. As discussed earlier in this chapter, BCIs for teams can enable novel forms of insight into collaborative dynamics, allowing users to reflect on interpersonal patterns that may otherwise go unnoticed. For example, as participants in my user-centered study noted, systems that surface shared attentional or emotional signals may help teams course-correct more quickly, promote transparency, and balance work more equitably. In ideal conditions, these systems can serve as supportive scaffolds for more empathetic and intentional collaboration.

However, the risks of surveillance, coercion, misinterpretation, and erosion of trust must also be taken seriously. This dissertation responds to those risks by proposing design guidelines to ensure BCI systems are deployed in ways that preserve user autonomy, data agency, and interpersonal dignity. Ethical use cases for BCIs, as surfaced in this work, are ones that emphasize user control and frame neural data as a lens for reflection rather than the sole determinant of performance, touching upon the Ethically Aligned Design (EAD) principles of transparency, data agency, and well-being [148].

These design choices are not just technical—they have the potential to influence how BCIs might shape workplace culture, interpersonal norms, and even the nature of team decision-making. Making explicit how ethical values map onto concrete system features provides a roadmap for deploying systems that enhance human capability without compromising ethical integrity.

8.5.3 Community and Public Engagement

This dissertation also contributes to broader engagement with the public and user communities around neurotechnology. In particular, it models and advocates for participatory, user-centered design processes in BCI development. As exemplified by our user-centered study, speculative design activities not only elicit concerns but also surface creative, value-driven visions for how technology might be used well. This participatory approach is grounded in the belief that systems should reflect the lived realities and aspirations of the people who will use them, not just the assumptions or wishes of researchers or designers.

By prioritizing accessibility and usability, particularly through tools that allow novice users to explore their own brain data, this work helps foster a culture of curiosity and reflection. Public-facing tools like BrainEx may help demystify neuroscience and empower individuals to think critically about cognitive processes and brain-sensing, which can serve to increase public understanding of and engagement with science and technology. In the long term, democratizing access to neurotechnology tools may also promote grassroots innovation and education, enabling members of the public to explore cognitive and emotional processes, understand the ethical stakes, and develop their own novel applications.

8.5.4 Evaluating Broader Impacts

Several pathways exist for evaluating the broader impacts proposed here. For reproducibility, future work could examine the uptake of tools and reporting practices proposed in this dissertation by tracking citations, replication studies, and adoption in research pipelines. Surveys or interviews with researchers or members of the public working with brain data could also gauge the usability and influence of frameworks like BrainEx or the reporting guidelines developed.

For collaborative BCI systems, longitudinal deployments in team-based workplace or educational settings could be used to assess perceived utility, ethical comfort, and outcomes on team dynamics. Additional user studies could explore whether the ethical safeguards proposed—such as consent controls or team-level aggregation—successfully mitigate the concerns participants raised in our work. Tracking public discourse or media representation around BCIs may also reveal changing perceptions and values.

Further evaluation might involve studying the spread and salience of team-facing BCIs in real-world use. This could include a meta-review of papers describing workplace BCI deployments, interviews with developers and users of such systems, and surveys gauging organizational attitudes toward brain sensing technologies. Such research would provide empirical grounding for assessing the ethical, social, and practical consequences of this dissertation's proposed approaches.

8.6 Conclusion

As BCIs move from clinical settings into team-oriented workplace environments, their ethical, technical, and social implications demand critical attention. This chapter has examined these implications through the lens of stakeholder concerns, ethical principles, and broader societal relevance. Drawing on a rich body of prior work and user-centered design research, I have argued that ethical BCI systems must prioritize individual agency, data stewardship, transparency, and interpersonal well-being. When designed with these values in mind, BCIs can serve as reflective tools for enhancing collaboration and self-understanding, rather than instruments of control or surveillance.

In parallel, this work advances reproducibility and accessibility in HCI's engagement with brain signal research, opening doors for cross-disciplinary innovation and public engagement. Through tools like BrainEx and a commitment to open, user-friendly design practices, this dissertation contributes to a future where neurotechnology is not only more ethical, but also more inclusive and comprehensible. Ultimately, these contributions support a vision for BCIs that enable thoughtful, responsible augmentation of team dynamics by centering people, their relationships, and their values at the heart of design.

Chapter 9

Concluding Remarks

This dissertation explored how brain-computer interfaces (BCIs) can be made more accessible, reproducible, and ethically integrated into collaborative settings. Through a series of technical contributions, empirical studies, and human-centered design interviews, this work addressed key challenges in the development and deployment of BCI systems: the lack of standard practices for reproducibility in brain signal research, the steep learning curve for non-experts hoping to work with neurophysiological data, and the ethical implications of using BCIs to support team dynamics and collaboration.

Chapters 3 and 4 addressed the first of these challenges, focusing on supporting open and transparent research practices. Chapter 3 presented a taxonomy of HCI research using brain signals, developed through a systematic review of 110 publications, and proposed an experiment model for more consistent and informative reporting. The model not only supports reproducibility and cross-disciplinary understanding, but also sheds light on the diversity of approaches and reporting strategies in BCI research. Chapter 4 introduced *BrainEx*, a visual exploration tool designed to make brain signal analysis more accessible to researchers without deep technical expertise. Combining high-performance similarity search and interactive data exploration, *BrainEx* supports new forms of insight generation, encouraging broader participation in BCI development. Together, these chapters contribute practical tools and frameworks that make it easier for others to engage with and build upon BCI research.

Chapters 7 and 6 shifted focus to the second core aim of the dissertation: understanding how brain signals can be leveraged in real-world collaborative settings to support more effective and meaningful team interactions.

Chapter 6 presented an exploratory empirical study examining the relationship between neural activity, team processes, and collaborative performance. Participants worked in small groups on a creative task while wearing EEG headsets. Analyses revealed modest but meaningful relationships between brain-based measures of engagement, workload, and synchrony, and behavioral indicators of team function. These findings support the technical feasibility of real-time BCI systems that help teams reflect on their own dynamics. While more research is needed to translate these signals into actionable feedback, this study lays groundwork for understanding the neurocognitive underpinnings of collaboration.

Chapter 7 complemented this exploration of practical feasibility with a user-centered design study investigating how team members and leaders imagine future BCI support tools, uncovering unmet needs, ethical concerns, and desired functionalities. Participants described the challenges of collaboration (such as incompatible team member personalities, motivation gaps, and miscommunication), as well as their hopes for technologies that could enhance mutual understanding, accountability, and synergy across the team. Informed by these insights, the chapter proposed a set of ethical guidelines for designing BCIs that center user agency, transparency, and well-being.

Beyond its technical and empirical contributions, this dissertation speaks to broader questions about the future of neurotechnology and the values that should guide its development. Chapter 8 engaged with the ethical and societal implications of deploying BCIs in team-oriented contexts. Drawing on prior literature, empirical data, and theoretical frameworks, it explored key considerations regarding the ethical principles of autonomy, privacy, and interpersonal responsibility. This work not only highlighted the potential risks of surveillance, coercion, and misinterpretation, but also argued for the transformative potential of BCIs when designed to promote self-awareness, empathy, and supportive rather than punitive feedback.

The broader impacts of this work span research, practice, and public engagement. By advancing reproducibility and accessibility in BCI research, this dissertation contributes to a more robust and inclusive scientific ecosystem. By modeling participatory, user-centered design, it demonstrates how neurotechnology can be shaped by the people it aims to serve. And by developing tools and frameworks that support values such as transparency, well-being, and autonomy, it offers a roadmap for responsible innovation in brain-computer interaction.

In sum, this dissertation lays the conceptual, methodological, and ethical foundation for BCI systems that are technically sound, but also responsive to the lived realities of users, envisioning a future in which BCIs amplify human insight, support collaboration, and enable richer relationships between people and technology. Future work may continue to build on these contributions by deploying functional BCI systems for teams, refining their feedback mechanisms, and evaluating their long-term impacts in workplace and educational environments. Despite considerable remaining challenges, the promise of designing neurotechnology in service of human values remains a powerful and worthwhile pursuit.

Bibliography

[1] National Research Council (US) Committee on Military and Intelligence Methodology for Emergent Neruophysiological and Cognitive/Neural Research in the Next Two Decades . 2008. *Emerging Cognitive Neuroscience and Related Technologies*. National Academies Press. Google-Books-ID: K6icAgAAQBAJ.

[2] John Aach and George M Church. 2001. Aligning gene expression time series with time warping algorithms. *Bioinformatics* 17, 6 (2001), 495–508.

[3] Daniel Afergan, Samuel W Hincks, Tomoki Shibata, and Robert JK Jacob. 2015. Phylter: a system for modulating notifications in wearables using physiological sensing. In *International conference on augmented cognition*. Springer, 167–177.

[4] Daniel Afergan, Evan M. Peck, Erin T. Solovey, Andrew Jenkins, Samuel W. Hincks, Eli T. Brown, Remco Chang, and Robert J.K. Jacob. 2014. Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3797–3806. `https://doi.org/10.1145/2556288.2557230`

[5] Daniel Afergan, Tomoki Shibata, Samuel W. Hincks, Evan M. Peck, Beste F. Yuksel, Remco Chang, and Robert J.K. Jacob. 2014. Brain-based target expansion. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 583–593. `https://doi.org/10.1145/2642918.2647414`

[6] Luís Aguiar-Conraria and Maria Joana Soares. 2014. The Continuous Wavelet Transform: Moving Beyond Uni- and Bivariate Analysis. *Journal of Economic Surveys* 28, 2 (2014), 344–375. `https://doi.org/10.1111/joes.12012`

[7] Audrey Aldridge, Eli Barnes, Cindy L. Bethel, Daniel W. Carruth, Marianna Kocturova, Matus Pleva, and Jozef Juhar. 2019. Accessible Electroencephalograms (EEGs): A Comparative Review with OpenBCI's Ultracortex Mark IV Headset. In *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*. 1–6. `https://doi.org/10.1109/RADIOELEK.2019.8733482`

[8] Teresa M. Amabile and Michael G. Pratt. 2016. The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior* 36 (Jan. 2016), 157–183. `https://doi.org/10.1016/j.riob.2016.10.001`

[9] Marvin Andujar, Chris S. Crawford, Anton Nijholt, France Jackson, and Juan E. Gilbert. 2015. Artistic brain-computer interfaces: the expression and stimulation of the user's affective state. *Brain-Computer Interfaces* 2, 2-3 (April 2015), 60–69. https://doi.org/10.1080/2326263X.2015.1104613

[10] Alissa N. Antle, Leslie Chesick, and Elgin-Skye Mclaren. 2018. Opening up the Design Space of Neurofeedback Brain–Computer Interfaces for Children. *ACM Transactions on Computer-Human Interaction* 24, 6 (Jan. 2018), 38:1–38:33. https://doi.org/10.1145/3131607

[11] Elizabeth M. Argyle, Adrian Marinescu, Max L. Wilson, Glyn Lawson, and Sarah Sharples. 2021. Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. *International Journal of Human-Computer Studies* 145 (2021), 102522. https://doi.org/10.1016/j.ijhcs.2020.102522

[12] P Aricò, G Borghini, G Di Flumeri, N Sciaraffa, and F Babiloni. 2018. Passive BCI beyond the lab: current trends and future directions. *Physiological measurement* 39, 8 (2018), 08TR02.

[13] Umer Asgher, Khurram Khalil, Muhammad Jawad Khan, Riaz Ahmad, Shahid Ikramullah Butt, Yasar Ayaz, Noman Naseer, and Salman Nazir. 2020. Enhanced Accuracy for Multiclass Mental Workload Detection Using Long Short-Term Memory for Brain–Computer Interface. *Frontiers in Neuroscience* 14 (2020). https://doi.org/10.3389/fnins.2020.00584

[14] Jonas Auda, Roman Heger, Thomas Kosch, Uwe Gruenefeld, and Stefan Schneegass. 2020. EasyEG: A 3D-printable Brain-Computer Interface. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 70–72. https://doi.org/10.1145/3379350.3416189

[15] Kregg Aytes. 1995. Comparing collaborative drawing tools and whiteboards: An analysis of the group process. *Computer Supported Cooperative Work (CSCW)* 4, 1 (March 1995), 51–71. https://doi.org/10.1007/BF00823363

[16] Ebrahim Babaei, Benjamin Tag, Tilman Dingler, and Eduardo Velloso. 2021. A Critique of Electrodermal Activity Practices at CHI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[17] Nicholas A Badcock, Petroula Mousikou, Yatin Mahajan, Peter De Lissa, Johnson Thie, and Genevieve McArthur. 2013. Validation of the Emotiv EPOC® EEG gaming system for measuring research quality auditory ERPs. *PeerJ* 1 (2013), e38.

[18] Lain Bailey. [n. d.]. OBS Studio. https://obsproject.com/

[19] Max Bain, Jaesung Huh, Tengda Han, and Andrew Zisserman. 2023. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. In *Proc. Interspeech 2023*. 4489–4493. https://www.isca-archive.org/interspeech_2023/bain23_interspeech.html

[20] Per Bak. 1996. *How Nature Works*. Springer, New York, NY. https://doi.org/10.1007/978-1-4757-5426-1

[21] Danushka Bandara, Senem Velipasalar, Sarah Bratt, and Leanne Hirshfield. 2018. Building predictive models of emotion with functional near-infrared spectroscopy. *International Journal of Human-Computer Studies* 110 (2018), 75 – 85. https://doi.org/10.1016/j.ijhcs.2017.10.001

[22] Patti Bao, Elizabeth Gerber, Darren Gergle, and David Hoffman. 2010. Momentum: getting and staying on topic during a brainstorm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1233–1236. https://doi.org/10.1145/1753326.1753511

[23] Suzanne T. Bell, Shanique G. Brown, Anthony Colaneri, and Neal Outland. 2018. Team composition and the ABCs of teamwork. *American Psychologist* 73, 4 (2018), 349–362. https://doi.org/10.1037/amp0000305

[24] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–20.

[25] Chris Berka, Daniel Levendowski, Michelle Lumicao, Alan Yau, Gene Davis, Tristan Zivkovic, Richard Olmstead, Patrice Tremoulet, and Patrick Craven. 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine* 78 (06 2007), B231–44.

[26] Guillermo Bernal, Tao Yang, Abhinandan Jain, and Pattie Maes. 2018. PhysioHMD: a conformable, modular toolkit for collecting physiological data from head-mounted displays. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 160–167. https://doi.org/10.1145/3267242.3267268

[27] Donald J. Berndt and James Clifford. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining* (Seattle, WA) *(AAAIWS'94)*. AAAI Press, 359–370.

[28] Akshay Bhagwatwar, Anne Massey, and Alan R. Dennis. 2013. Creative Virtual Environments: Effect of Supraliminal Priming on Team Brainstorming. In *2013 46th Hawaii International Conference on System Sciences*. 215–224. https://doi.org/10.1109/HICSS.2013.152 ISSN: 1530-1605.

[29] Maneesh Bilalpur, Mohan Kankanhalli, Stefan Winkler, and Ramanathan Subramanian. 2018. EEG-based Evaluation of Cognitive Workload Induced by Acoustic Parameters for Data Sonification. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (Boulder, CO, USA) *(ICMI '18)*. ACM, New York, NY, USA, 315–323. https://doi.org/10.1145/3242969.3243016

[30] Philip Bobko, Leanne Hirshfield, Lucca Eloy, Cara Spencer, Emily Doherty, Jack Driscoll, and Hannah Obolsky. 2022. Human-agent teaming and trust calibration: a theoretical framework, configurable testbed, empirical illustration, and implications for the development of adaptive systems. *Theoretical Issues in Ergonomics Science* 0, 0 (June 2022), 1–25. https://doi.org/10.1080/1463922X.2022.2086644

[31] Marietjie Wilhelmina Maria Botes. 2022. Brain Computer Interfaces and Human Rights: Brave new rights for a brave new world. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1154–1161. https://doi.org/10.1145/3531146.3533176

[32] L. Botrel, E.M. Holz, and A. Kübler. 2015. Brain Painting V2: evaluation of P300-based brain-computer interface for creative expression by an end-user following the user-centered design. *Brain-Computer Interfaces* 2, 2-3 (April 2015), 135–149. https://doi.org/10.1080/2326263X.2015.1100038

[33] Elisabeth Brauner, Margarete Boos, and Michaela Kolbe. 2018. *The Cambridge Handbook of Group Interaction Analysis*. Cambridge University Press. Google-Books-ID: KnNqDwAAQBAJ.

[34] Todd S Braver. 2012. The variable nature of cognitive control: a dual mechanisms framework. *Trends in cognitive sciences* 16, 2 (2012), 106–113.

[35] Anne-Marie Brouwer, Thorsten O Zander, Jan BF Van Erp, Johannes E Korteling, and Adelbert W Bronkhorst. 2015. Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience* 9 (2015), 136.

[36] Melanie S. Brucks and Jonathan Levav. 2022. Virtual communication curbs creative idea generation. *Nature* 605, 7908 (May 2022), 108–112. https://doi.org/10.1038/s41586-022-04643-y

[37] Clemens Brunner, Niels Birbaumer, Benjamin Blankertz, Christoph Guger, Andrea Kübler, Donatella Mattia, José del R Millán, Felip Miralles, Anton Nijholt, Eloy Opisso, et al. 2015. BNCI Horizon 2020: towards a roadmap for the BCI community. *Brain-computer interfaces* 2, 1 (2015), 1–10.

[38] Paulo Buono and Adalberto Lafcadio Simeone. 2008. Interactive shape specification for pattern search in time series. *AVI '08: Proceedings of the working conference on Advanced visual interfaces* (2008), 480–481.

[39] Christopher G Burns and Stephen H Fairclough. 2015. Use of auditory event-related potentials to measure immersion during a computer game. *International Journal of Human-Computer Studies* 73 (2015), 107–114. https://doi.org/10.1016/j.ijhcs.2014.09.002

[40] Sasha Burwell, Matthew Sample, and Eric Racine. 2017. Ethical aspects of brain computer interfaces: a scoping review. *BMC medical ethics* 18, 1 (2017), 1–11.

[41] Enrico Gianluca Caiani, A Porta, Giuseppe Baselli, M Turiel, S Muzzupappa, F Pieruzzi, C Crema, A Malliani, and Sergio Cerutti. 1998. Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. In *Computers in Cardiology 1998. Vol. 25 (Cat. No. 98CH36292)*. IEEE, 73–76.

[42] Francisco M Calisto, Alfredo Ferreira, Jacinto C Nascimento, and Daniel Gonçalves. 2017. Towards touch-based medical image diagnosis annotation. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 390–395.

[43] Francisco Maria Calisto, Nuno Nunes, and Jacinto C Nascimento. 2020. Breastscreening: On the use of multi-modality in medical imaging diagnosis. In *Proceedings of the International Conference on Advanced Visual Interfaces*. 1–5.

[44] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2021. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human-Computer Studies* 150 (2021), 102607.

[45] Cédric Cannard, Helané Wahbeh, and Arnaud Delorme. 2021. Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 3603–3610.

[46] Sung-Hyuk Cha. 2007. Comprehensive survey on distance/similarity measures between probability density functions. *City* 1, 2 (2007), 1.

[47] Joel Chan, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T. Solovey, Krzysztof Z. Gajos, and Steven P. Dow. 2017. Semantically Far Inspirations Considered Harmful? Accounting for Cognitive States in Collaborative Ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)*. Association for Computing Machinery, New York, NY, USA, 93–105. https://doi.org/10.1145/3059454.3059455

[48] Britton Chance, Endla Anday, Shoko Nioka, Shuoming Zhou, Long Hong, Katherine Worden, C Li, T Murray, Y Ovetsky, D Pidikiti, et al. 1998. A novel method for fast imaging of brain function, non-invasively, with light. *Optics express* 2, 10 (1998), 411–423.

[49] Guillaume Chanel, J. Matias Kivikangas, and Niklas Ravaja. 2012. Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interacting with Computers* 24, 4 (July 2012), 306–316. https://doi.org/10.1016/j.intcom.2012.04.012

[50] Guillaume Chanel and Christian Mühl. 2015. Connecting Brains and Bodies: Applying Physiological Computing to Support Social Interaction. *Interacting with Computers* 27, 5 (Sept. 2015), 534–550. https://doi.org/10.1093/iwc/iwv013

[51] WA Chaovalitwongse and PM Pardalos. 2008. On the time series support vector machine using dynamic time warping kernel for brain activity classification. *Cybernetics and Systems Analysis* 44, 1 (2008), 125–138.

[52] Max Chen, Erin Solovey, and Gillian Smith. 2023. Impact of BCI-Informed Visual Effect Adaptation in a Walking Simulator. In *Proceedings of the 18th International Conference on the Foundations of Digital Games*. 1–8.

[53] Fu-Yin Cherng, Yi-Chen Lee, Jung-Tai King, and Wen-Chieh Lin. 2019. Measuring the Influences of Musical Parameters on Cognitive and Behavioral Responses to Audio Notifications Using EEG and Large-scale Online Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 409, 12 pages. https://doi.org/10.1145/3290605.3300639

[54] Prerna Chikersal, Maria Tomprou, Young Ji Kim, Anita Williams Woolley, and Laura Dabbish. 2017. Deep Structures of Collaboration: Physiological Correlates of Collective Intelligence and Group Satisfaction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 873–888. https://doi.org/10.1145/2998181.2998250

[55] Sanjib Chowdhury. 2005. The Role of Affect- and Cognition-based Trust in Complex Knowledge Sharing. *Journal of Managerial Issues* 17, 3 (2005), 310–326. https://www.jstor.org/stable/40604504

[56] S. Mealla Cincuegrani, S. Jordà, and A. Väljamäe. 2016. Physiopucks: Increasing User Motivation by Combining Tangible and Implicit Physiological Interaction. *ACM Transactions on Computer-Human Interaction* 23, 1 (Feb. 2016), 4:1–4:22. https://doi.org/10.1145/2838732

[57] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.

[58] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. HARK No More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 141:1–141:12. https://doi.org/10.1145/3173574.3173715

[59] Moreno I. Coco and Rick Dale. 2014. Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in Psychology* 5 (2014). https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00510

[60] Allen Coin, Megan Mulder, and Veljko Dubljević. 2020. Ethical Aspects of BCI Technology: What Is the State of the Art? *Philosophies* 5, 4 (2020), 31.

[61] Chris S. Crawford and Juan E. Gilbert. 2019. Brains and Blocks: Introducing Novice Programmers to Brain-Computer Interface Application Development. *ACM Transactions on Computing Education* 19, 4 (July 2019), 39:1–39:27. `https://doi.org/10.1145/3335815`

[62] Igor Crk, Timothy Kluthe, and Andreas Stefik. 2015. Understanding Programming Expertise: An Empirical Study of Phasic Brain Wave Changes. *ACM Trans. Comput.-Hum. Interact.* 23, 1, Article 2 (Dec. 2015), 29 pages. `https://doi.org/10.1145/2829945`

[63] Edward Cutrell and Desney Tan. 2008. BCI for passive input in HCI. In *Proceedings of CHI 2008 Workshop on Brain-Computer Interfaces for HCI and Games*. 1–3.

[64] Artur Czeszumski, Sophie Hsin-Yi Liang, Suzanne Dikker, Peter König, Chin-Pang Lee, Sander L. Koole, and Brent Kelsen. 2022. Cooperative Behavior Evokes Interbrain Synchrony in the Prefrontal and Temporoparietal Cortex: A Systematic Review and Meta-Analysis of fNIRS Hyperscanning Studies. *eNeuro* 9, 2 (April 2022), ENEURO.0268–21.2022. `https://doi.org/10.1523/ENEURO.0268-21.2022`

[65] Ian Daly, Duncan Williams, Alexis Kirke, James Weaver, Asad Malik, Faustina Hwang, Eduardo Miranda, and Slawomir J. Nasuto. 2016. Affective brain–computer music interfacing. *Journal of Neural Engineering* 13, 4 (July 2016), 046022. `https://doi.org/10.1088/1741-2560/13/4/046022`

[66] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Annh Ratanamahatana, and Eamonn Keogh. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica* 6, 6 (2019), 1293–1305.

[67] Luciënne A. de With, Nattapong Thammasan, and Mannes Poel. 2022. Detecting Fear of Heights Response to a Virtual Reality Environment Using Functional Near-Infrared Spectroscopy. *Frontiers in Computer Science* 3 (2022). `https://doi.org/10.3389/fcomp.2021.652550`

[68] Arwen H. DeCostanza, Amar R. Marathe, Addison Bohannon, A. William Evans, Edward T. Palazzolo, Jason S. Metcalfe, and Kaleb McDowell. 2018. *Enhancing humanagent teaming with individualized, adaptive technologies: A discussion of critical scientific questions*. Technical Report. US Army Research Laboratory Aberdeen Proving Ground United States.

[69] Arnaud Delorme. 2023. EEG is better left alone. *Scientific Reports* 13, 1 (Feb. 2023), 1–12. `https://doi.org/10.1038/s41598-023-27528-0`

[70] Arnaud Delorme and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134, 1 (2004), 9–21.

[71] Laura Devendorf, Joanne Lo, Noura Howell, Jung Lin Lee, Nan-Wei Gong, M. Emre Karagozler, Shiho Fukuhara, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. "I don't Want to Wear a Screen": Probing Perceptions of and Possibilities for Dynamic Displays on Clothing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 6028–6039. https://doi.org/10.1145/2858036.2858192

[72] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. 2008. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. *Proc. VLDB Endow.* 1, 2 (Aug. 2008), 1542–1552. https://doi.org/10.14778/1454159.1454226

[73] Katherine L Dixon-Gordon, Samantha L Bernecker, and Kara Christensen. 2015. Recent innovations in the field of interpersonal emotion regulation. *Current Opinion in Psychology* 3 (June 2015), 36–42. https://doi.org/10.1016/j.copsyc.2015.02.001

[74] Sidney D'Mello, Kristopher Kopp, Robert Earl Bixler, and Nigel Bosch. 2016. Attending to Attention: Detecting and Combating Mind Wandering during Computerized Reading. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1661–1669. https://doi.org/10.1145/2851581.2892329

[75] Silke Dodel, Joseph Cohn, Jochen Mersmann, Phan Luu, Chris Forsythe, and Viktor Jirsa. 2011. Brain Signatures of Team Performance. In *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems (Lecture Notes in Computer Science)*, Dylan D. Schmorrow and Cali M. Fidopiastis (Eds.). Springer, Berlin, Heidelberg, 288–297. https://doi.org/10.1007/978-3-642-21852-1_35

[76] Eamon Doherty, Gilbert Cockton, Chris Bloor, and Dennis Benigno. 2001. Improving the Performance of the Cyberlink Mental Interface with "Yes / No Program". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. ACM, New York, NY, USA, 69–76. https://doi.org/10.1145/365024.365038

[77] Lukas Dommer, Natalie Jäger, Felix Scholkmann, Martin Wolf, and Lisa Holper. 2012. Between-brain coherence during joint n-back task performance: A two-person functional near-infrared spectroscopy study. *Behavioural Brain Research* 234, 2 (Oct. 2012), 212–222. https://doi.org/10.1016/j.bbr.2012.06.024

[78] Gesine Dreisbach and Kerstin Fröber. 2019. On How to Be Flexible (or Not): Modulation of the Stability-Flexibility Balance. *Current Directions in Psychological Science* 28, 1 (Feb. 2019), 3–9. https://doi.org/10.1177/0963721418800030

[79] James E. Driskell, Gerald F. Goodwin, Eduardo Salas, and Patrick Gavan O'Shea. 2006. What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice* 10 (2006), 249–271. https://doi.org/10.1037/1089-2699.10.4.249

[80] Jayesh Dubey, Mihin Sumaria, Erden Oktay, Yu Li, Ziheng Li, Rodica Neamtu, and Erin T Solovey. 2019. Towards Neuroadaptive Technology Using Time Warped Distances for Similarity Exploration of Brain Data. In *The Second Neuroadaptive Technology Conference*. 48.

[81] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3170427.3188395

[82] J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle. 1987. Recurrence Plots of Dynamical Systems. *Europhysics Letters* 4, 9 (Nov. 1987), 973. https://doi.org/10.1209/0295-5075/4/9/004

[83] Ernest Edmonds, Dave Everitt, Michael Macaulay, and Greg Turner. 2004. On physiological computing with an application in interactive art. *Interacting with Computers* 16, 5 (Oct. 2004), 897–915. https://doi.org/10.1016/j.intcom.2004.08.003

[84] Stefan K. Ehrlich, Kat R. Agres, Cuntai Guan, and Gordon Cheng. 2019. A closed-loop, music-based brain-computer interface for emotion mediation. *PLOS ONE* 14, 3 (March 2019), e0213516. https://doi.org/10.1371/journal.pone.0213516

[85] Lucca Eloy, Cara Spencer, Emily Doherty, and Leanne Hirshfield. 2023. Capturing the Dynamics of Trust and Team Processes in Human-Human-Agent Teams via Multidimensional Neural Recurrence Analyses. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 122:1–122:23. https://doi.org/10.1145/3579598

[86] Andéol Evain, Ferran Argelaguet, Nicolas Roussel, Géry Casiez, and Anatole Lécuyer. 2017. Can I Think of Something Else when Using a BCI?: Cognitive Demand of an SSVEP-based BCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5120–5125. https://doi.org/10.1145/3025453.3026037

[87] Stephen Fairclough. 2014. Physiological data must remain confidential. *Nature News* 505, 7483 (2014), 263.

[88] Stephen Fairclough. 2015. A Closed-Loop Perspective on Symbiotic Human-Computer Interaction. In *Symbiotic Interaction (Lecture Notes in Computer Science)*, Benjamin Blankertz, Giulio Jacucci, Luciano Gamberini, Anna Spagnolli, and Jonathan Freeman (Eds.). Springer International Publishing, Cham, 57–67. https://doi.org/10.1007/978-3-319-24917-9_6

[89] Stephen Fairclough and Kiel Gilleade. 2012. Construction of the biocybernetic loop: a case study. In *Proceedings of the 14th ACM international conference on Multimodal interaction (ICMI '12)*. Association for Computing Machinery, New York, NY, USA, 571–578. https://doi.org/10.1145/2388676.2388797

[90] Stephen H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with Computers* 21, 1 (Jan. 2009), 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

[91] Stephen H. Fairclough. 2022. Chapter 1 - Designing human-computer interaction with neuroadaptive technology. In *Current Research in Neuroadaptive Technology*, Stephen H. Fairclough and Thorsten O. Zander (Eds.). Academic Press, 1–15. https://doi.org/10.1016/B978-0-12-821413-8.00006-3

[92] Christos Faloutsos, Mudumbai Ranganathan, and Yannis Manolopoulos. 1994. Fast subsequence matching in time-series databases. *Acm Sigmod Record* 23, 2 (1994), 419–429.

[93] Adam R Ferguson, Jessica L Nielson, Melissa H Cragin, Anita E Bandrowski, and Maryann E Martone. 2014. Big data from small data: data-sharing in the'long tail'of neuroscience. *Nature neuroscience* 17, 11 (2014), 1442–1447.

[94] M Ferrari and V Quaresima. 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage* 63 (2012), 921–935.

[95] Stéphanie Fleck and Martin Hachet. 2016. Making Tangible the Intangible: Hybridization of the Real and the Virtual to Enhance Learning of Abstract Phenomena. *Frontiers in ICT* 3 (Dec. 2016). https://doi.org/10.3389/fict.2016.00030

[96] Cesar Augusto Fontanillo Lopez, Guangye Li, and Dingguo Zhang. 2020. Beyond Technologies of Electroencephalography-Based Brain-Computer Interfaces: A Systematic Review From Commercial and Ethical Aspects. *Frontiers in Neuroscience* 14 (Dec. 2020). https://doi.org/10.3389/fnins.2020.611130

[97] Donelson R. Forsyth. 2018. *Group Dynamics*. Cengage Learning. Google-Books-ID: vg9EDwAAQBAJ.

[98] Frederick Freeman, Peter Mikulka, Mark Scerbo, Lawrence Prinzel, and Keith Clouatre. 2000. Evaluation of a Psychophysiologically Controlled Adaptive Automation System, Using Performance on a Tracking Task. *Applied psychophysiology and biofeedback* 25 (07 2000), 103–15. https://doi.org/10.1023/A:1009566809021

[99] Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. 2016. Framework for Electroencephalography-based Evaluation of User Experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 2283–2294. https://doi.org/10.1145/2858036.2858525

[100] Jérémy Frey, Renaud Gervais, Stéphanie Fleck, Fabien Lotte, and Martin Hachet. 2014. Teegi: tangible EEG interface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)*. Association for Computing Machinery, New York, NY, USA, 301–308. https://doi.org/10.1145/2642918.2647368

[101] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3290605.3300619

[102] A. Gabriel, D. Monticolo, M. Camargo, and M. Bourgault. 2016. Creativity support systems: A systematic mapping study. *Thinking Skills and Creativity* 21 (Sept. 2016), 109–122. https://doi.org/10.1016/j.tsc.2016.05.009

[103] Weihao Gao, Sewoong Oh, and Pramod Viswanath. 2018. Demystifying Fixed k - Nearest Neighbor Information Estimators. *IEEE Transactions on Information Theory* 64, 8 (Aug. 2018), 5629–5661. https://doi.org/10.1109/TIT.2018.2807481

[104] Fabiola H. Gerpott, Nale Lehmann-Willenbrock, Jeroen D. Silvis, and Mark Van Vugt. 2018. In the eye of the beholder? An eye-tracking experiment on emergent leadership in team interactions. *The Leadership Quarterly* 29, 4 (Aug. 2018), 523–532. https://doi.org/10.1016/j.leaqua.2017.11.003

[105] Alan Gevins and Michael Smith. 2003. Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interaction. Theor. Issues Ergonomics Sci. 4(1-2), 113-131. *Theoretical Issues in Ergonomics Science* 4 (01 2003), 113–131. https://doi.org/10.1080/14639220210159717

[106] Georgios Giasemidis and Danica Vukadinović Greetham. 2018. Optimising Parameters in Recurrence Quantification Analysis of Smart Energy Systems. In *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*. 1–7. https://doi.org/10.1109/IISA.2018.8633648

[107] Christiane Glatz, Stas S. Krupenia, Heinrich H. Bülthoff, and Lewis L. Chuang. 2018. Use the Right Sound for the Right Job: Verbal Commands and Auditory Icons for a Task-Management System Favor Different Information Processes in the Brain. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 472, 13 pages. https://doi.org/10.1145/3173574.3174046

[108] Anna Gogolou, Theophanis Tsandilas, Themis Palpanas, and Anastasia Bezerianos. 2019. Comparing Similarity Perception in Time Series Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25 (2019), 523–533.

[109] Emma C. Gordon and Anil K. Seth. 2024. Ethical considerations for the use of brain–computer interfaces for cognitive enhancement. *PLOS Biology* 22, 10 (Oct. 2024), e3002899. https://doi.org/10.1371/journal.pbio.3002899

[110] Jamie C. Gorman, Nancy J. Cooke, Polemnia G. Amazeen, and Shannon Fouse. 2012. Measuring Patterns in Team Interaction Sequences Using a Discrete Recurrence Approach. *Human Factors* 54, 4 (Aug. 2012), 503–517. https://doi.org/10.1177/0018720811426140

[111] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. 2013. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience* 7, 267 (2013), 1–13. https://doi.org/10.3389/fnins.2013.00267

[112] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. 2014. MNE software for processing MEG and EEG data. *Neuroimage* 86 (2014), 446–460.

[113] A. Grinsted, J. C. Moore, and S. Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* 11, 5/6 (Nov. 2004), 561–566. https://doi.org/10.5194/npg-11-561-2004

[114] g.tec medical engineering GmbH. 2024. Unicorn Hybrid Black. https://www.gtec.at/product/unicorn-hybrid-black/

[115] Stephen J. Guastello. 2016. Physiological Synchronization in a Vigilance Dual Task. *Nonlinear Dynamics, Psychology, and Life Sciences* 20, 1 (Jan. 2016), 49–80.

[116] Stephen J. Guastello and Anthony F. Peressini. 2021. A Comparison of Four Dyadic Synchronization Models. *Nonlinear Dynamics, Psychology, and Life Sciences* 25, 1 (Jan. 2021), 19–39.

[117] Stephen J Guastello and Anthony F Peressini. 2023. Quantifying synchronization in groups with three or more members using SyncCalc: The driver-empath model of group dynamics. *Group Dynamics: Theory, Research, and Practice* (2023).

[118] C. Guger, A. Schlogl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller. 2001. Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9, 1 (March 2001), 49–58. https://doi.org/10.1109/7333.918276

[119] J. P. Guilford. 1950. Creativity. *American Psychologist* 5 (1950), 444–454. https://doi.org/10.1037/h0063487

[120] Janine Viol Hacker, Michael Johnson, Carol Saunders, and Amanda L. Thayer. 2019. Trust in Virtual Teams: A Multidisciplinary Review and Integration. *Australasian Journal of Information Systems* 23 (Jan. 2019). https://doi.org/10.3127/ajis.v23i0.1757

[121] Aamir Saeed Malik Hafeez Ullah Amin. 2017. *Designing EEG Experiments For Studying The Brain.* Elsevier. http://archive.org/details/2017-malik-designing-eeg-experiments-for-studying-the-brain

[122] U Hakim, S De Felice, P Pinti, X Zhang, J.A. Noah, Y Ono, P. W. Burgess, A Hamilton, J Hirsch, and I Tachtsidis. 2023. Quantification of inter-brain coupling: A review of

current methods used in haemodynamic and electrophysiological hyperscanning studies. *NeuroImage* 280 (Oct. 2023), 120354. https://doi.org/10.1016/j.neuroimage.2023.120354

[123] Peter Halling, Paul F. Fitzpatrick, Frank M. Raushel, Johann Rohwer, Santiago Schnell, Ulrike Wittig, Roland Wohlgemuth, and Carsten Kettner. [n. d.]. An Empirical Analysis of Enzyme Function Reporting for Experimental Reproducibility: Missing/Incomplete Information in Published Papers. 242 ([n. d.]), 22–27. https://doi.org/10.1016/j.bpc.2018.08.004

[124] Yuanning Han, Ziyi Qiu, Jiale Cheng, and RAY LC. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3613904.3642133

[125] Sarah Harvey. 2014. Creative Synthesis: Exploring the Process of Extraordinary Group Creativity. *Academy of Management Review* 39, 3 (July 2014), 324–343. https://doi.org/10.5465/amr.2012.0224

[126] Bryan Hernandez-Cuevas, William Egbert, Andre Denham, Ajay Mehul, and Chris S. Crawford. 2020. Changing Minds: Exploring Brain-Computer Interface Experiences with High School Students. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382981

[127] Lawrence J. Hettinger, Pedro Branco, L. Miguel Encarnacao, and Paolo Bonato. 2003. Neuroadaptive technologies: Applying neuroergonomics to the design of advanced interfaces. *Theoretical Issues in Ergonomics Science* (Jan. 2003). https://doi.org/10.1080/1463922021000020918

[128] Elisabeth Hildt. 2019. Multi-Person Brain-To-Brain Interfaces: Ethical Issues. *Frontiers in Neuroscience* 13 (Nov. 2019). https://doi.org/10.3389/fnins.2019.01177

[129] Johan Himberg, Aapo Hyvärinen, and Fabrizio Esposito. 2004. Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage* 22, 3 (2004), 1214–1222.

[130] H. Hochheiser and B. Shneiderman. 2004. Dynamic Query Tools for Time Series Data Sets: Timebox Widgets for Interactive Exploration. *Information Visualization* 3 (2004). Issue 1.

[131] L Hocke, I Oni, C Duszynski, A Corrigan, B Frederick, and J Dunn. 2018. Automated Processing of fNIRS Data—A Visual Guide to the Pitfalls and Consequences. Algorithms. *Algorithms* 11 (2018).

[132] Anu Holm, Kristian Lukander, Jussi Korpela, Mikael Sallinen, and Kiti Müller. 2009. Estimating Brain Load from the EEG. *TheScientificWorldJournal* 9 (02 2009), 639–51. https://doi.org/10.1100/tsw.2009.83

[133] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2017. Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. Association for Computing Machinery, New York, NY, USA, 257–266. https://doi.org/10.1145/3027385.3027451

[134] Karen Holtzblatt and Hugh Beyer. 1997. *Contextual design: defining customer-centered systems*. Elsevier.

[135] Bernhard Hommel. 2015. Chapter Two - Between Persistence and Flexibility: The Yin and Yang of Action Control. In *Advances in Motivation Science*, Andrew J. Elliot (Ed.). Vol. 2. Elsevier, 33–67. https://doi.org/10.1016/bs.adms.2015.04.003

[136] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough?: On the Extent and Content of Replications in Human-computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3523–3532. https://doi.org/10.1145/2556288.2557004

[137] Xin Hou, Zong Zhang, Chen Zhao, Lian Duan, Yilong Gong, Zheng Li, and Chaozhe Zhu. 2021. NIRS-KIT: a MATLAB toolbox for both resting-state and task fNIRS data analysis. *Neurophotonics* 8, 1 (January 2021), 010802. https://doi.org/10.1117/1.nph.8.1.010802

[138] Alicia Howell-Munson, Emily Doherty, Peter Gavriel, Claire Nicolas, Adam Norton, Rodica Neamtu, Holly Yanco, Yi-Ning Wu, and Erin T Solovey. 2022. Towards brain metrics for improving multi-agent adaptive human-robot collaboration: A preliminary study. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 1–10.

[139] Alicia Howell-Munson, Christopher Micek, Ziheng Li, Michael Clements, Andrew C. Nolan, Jackson Powell, Erin T. Solovey, and Rodica Neamtu. 2022. BrainEx: Interactive Visual Exploration and Discovery of Sequence Similarity in Brain Signals. *Proc. ACM Hum.-Comput. Interact.* 6, EICS (June 2022), 162:1–162:41. https://doi.org/10.1145/3534516

[140] Alicia Howell-Munson, Deniz Sonmez, Erin Walker, Catherine Arrington, and Erin T. Solovey. 2021. Preliminary steps towards detection of proactive and reactive control states during learning with fNIRS brain signals. In *Proceedings of the First International Workshop on Multimodal Artificial Intelligence in Education (MAIED 2021)*, Vol. 2902.

[141] Mo Hu, Tripp Shealy, Julie Milovanovic, and John Gero. 2022. Neurocognitive feedback: a prospective approach to sustain idea generation during design brainstorming.

International Journal of Design Creativity and Innovation 10, 1 (Jan. 2022), 31–50. https://doi.org/10.1080/21650349.2021.1976678

[142] Wei Hu and Bing Cheng Si. 2016. Technical note: Multiple wavelet coherence for untangling scale-specific and localized multivariate relationships in geosciences. *Hydrology and Earth System Sciences* 20, 8 (Aug. 2016), 3183–3191. https://doi.org/10.5194/hess-20-3183-2016

[143] Haiyun Huang, Qiuyou Xie, Jiahui Pan, Yanbin He, Zhenfu Wen, Ronghao Yu, and Yuanqing Li. 2021. An EEG-Based Brain Computer Interface for Emotion Recognition and Its Application in Patients with Disorder of Consciousness. *IEEE Transactions on Affective Computing* 12, 4 (Oct. 2021), 832–842. https://doi.org/10.1109/TAFFC.2019.2901456

[144] Zhe Huang, Liang Wang, Giles Blaney, Christopher Slaughter, Devon McKeon, Ziyu Zhou, Robert Jacob, and Michael Hughes. 2021. The Tufts fNIRS Mental Workload Dataset & Benchmark for Brain-Computer Interfaces that Generalize. *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks* 1 (Dec. 2021). https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/bd686fd640be98efaae0091fa301e613-Abstract-round2.html

[145] Edward C. K. Hung and Clifford S. T. Choy. 2013. Conceptual Recombination: A method for producing exploratory and transformational creativity in creative works. *Knowledge-Based Systems* 53 (Nov. 2013), 1–12. https://doi.org/10.1016/j.knosys.2013.07.007

[146] Theodore J Huppert, Solomon G Diamond, Maria A Franceschini, and David A Boas. 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics* 48, 10 (April 2009), D280—98. https://doi.org/10.1364/ao.48.00d280

[147] IEEE. [n. d.]. IEEE Neuroethics Framework. https://brain.ieee.org/publications/ieee-neuroethics-framework/

[148] IEEE. 2019. Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. *Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* (2019), 1–294.

[149] Marcello Ienca and Pim Haselager. 2016. Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity. *Ethics and Information Technology* 18, 2 (2016), 117–129.

[150] Marcello Ienca, Pim Haselager, and Ezekiel J Emanuel. 2018. Brain leaks and consumer neurotechnology. *Nature biotechnology* 36, 9 (2018), 805–810.

[151] A. Inselberg and B. Dimsdale. 1990. Parallel coordinates: a tool for visualizing multidimensional geometry. *Proceedings of the First IEEE Conference on Visualization: Visualization '90* (1990), 361–378.

[152] Lina Elsherif Ismail and Waldemar Karwowski. 2020. Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis. *PLOS ONE* 15, 12 (Dec. 2020), e0242857. https://doi.org/10.1371/journal.pone.0242857

[153] Giulio Jacucci, Stephen Fairclough, and Erin T. Solovey. 2015. Physiological Computing. *Computer* 48, 10 (Oct. 2015), 12–16. https://doi.org/10.1109/MC.2015.291

[154] Sirkka L. Jarvenpaa and Liisa Välikangas. 2020. Advanced Technology and End-Time in Organizations: A Doomsday for Collaborative Creativity? *Academy of Management Perspectives* 34, 4 (Nov. 2020), 566–584. https://doi.org/10.5465/amp.2019.0040

[155] Jan Jarvis, Felix Putze, Dominic Heger, and Tanja Schultz. 2011. Multimodal Person Independent Recognition of Workload Related Biosignal Patterns. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. ACM, New York, NY, USA, 205–208. https://doi.org/10.1145/2070481.2070516

[156] Patrick Jermann, Dejana Mullins, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2011. Collaborative Gaze Footprints: Correlates of Interaction Quality. (June 2011). https://repository.isls.org//handle/1/2446

[157] Camille Jeunet, Stefan Debener, Fabien Lotte, Jérémie Mattout, Reinhold Scherer, and Catharina Zich. 2018. Mind the Traps! Design Guidelines for Rigorous BCI Experiments. In *Brain–Computer Interfaces Handbook*. CRC Press.

[158] Weiwei Jiang, Zhanna Sarsenbayeva, Niels van Berkel, Chaofan Wang, Difeng Yu, Jing Wei, Jorge Goncalves, and Vassilis Kostakos. 2021. User Trust in Assisted Decision-Making Using Miniaturized Near-Infrared Spectroscopy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 153, 16 pages. https://doi.org/10.1145/3411764.3445710

[159] Daniel Johnson, Peta Wyeth, Madison Clark, and Christopher Watling. 2015. Cooperative Game Play with Avatars and Agents: Differences in Brain Activity and the Experience of Play. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3721–3730. https://doi.org/10.1145/2702123.2702468

[160] Andrew R. Jones, Michael Miller, Ruedi Aebersold, Rolf Apweiler, Catherine A. Ball, Alvis Brazma, James DeGreef, Nigel Hardy, Henning Hermjakob, Simon J. Hubbard, Peter Hussey, Mark Igra, Helen Jenkins, Randall K. Julian, Kent Laursen, Stephen G. Oliver, Norman W. Paton, Susanna-Assunta Sansone, Ugis Sarkans, Christian J. Stoeckert, Chris F. Taylor, Patricia L. Whetzel, Joseph A. White, Paul Spellman, and Angel Pizarro. 2007. The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nature Biotechnology* 25, 10 (Oct. 2007), 1127–1133. https://doi.org/10.1038/nbt1347

[161] Keith S Jones, Matthew Middendorf, Grant R McMillan, Gloria Calhoun, and Joel Warm. 2003. Comparing mouse and steady-state visual evoked response-based control. *Interacting with Computers* 15, 4 (2003), 603–621.

[162] Altyngul T. Kamzanova, Almira M. Kustubayeva, and Gerald Matthews. 2014. Use of EEG Workload Indices for Diagnostic Monitoring of Vigilance Decrement. *Human Factors* 56, 6 (2014), 1136–1149. https://doi.org/10.1177/0018720814526617 PMID: 25277022.

[163] Emre Karakoc, Artem Cherkasov, and S Cenk Sahinalp. 2006. Distance based algorithms for small biomolecule classification and structural similarity search. *Bioinformatics* 22, 14 (2006), e243–e251.

[164] Rohit J Kate. 2016. Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery* 30, 2 (2016), 283–312.

[165] Stamos Katsigiannis and Naeem Ramzan. 2018. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics* 22, 1 (Jan. 2018), 98–107. https://doi.org/10.1109/JBHI.2017.2688239

[166] Aleksandra Kawala-Sterniuk, Natalia Browarska, Amir Al-Bakri, Mariusz Pelc, Jaroslaw Zygarlicki, Michaela Sidikova, Radek Martinek, and Edward Jacek Gorzelanczyk. 2021. Summary of over Fifty Years with Brain-Computer Interfaces—A Review. *Brain Sciences* 11, 1 (Jan. 2021), 43. https://doi.org/10.3390/brainsci11010043

[167] Sadaf Kazi, Salar Khaleghzadegan, Julie V. Dinh, Mark J. Shelhamer, Adam Sapirstein, Lee A. Goeddel, Nnenna O. Chime, Eduardo Salas, and Michael A. Rosen. 2021. Team Physiological Dynamics: A Critical Review. *Human Factors* 63, 1 (Feb. 2021), 32–65. https://doi.org/10.1177/0018720819874160

[168] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems* 3, 3 (2001), 263–286.

[169] Eamonn Keogh and Chotirat Ann Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowledge and information systems* 7, 3 (2005), 358–386.

[170] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W. Malone. 2017. What Makes a Strong Team? Using Collective Intelligence to Predict Team Performance in League of Legends. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2316–2329. https://doi.org/10.1145/2998181.2998185

[171] Florian Klonek, Melissa Twemlow, Maria Tims, and Sharon K. Parker. 2024. It's About Time! Understanding the Dynamic Team Process-Performance Relationship

Using Micro- and Macroscale Time Lenses. *Group & Organization Management* (Sept. 2024), 10596011241278556. https://doi.org/10.1177/10596011241278556

[172] Silvia Erika Kober and Christa Neuper. 2012. Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies* 70, 9 (Sept. 2012), 577–587. https://doi.org/10.1016/j.ijhcs.2012.03.004

[173] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Transactions on Affective Computing* 3, 1 (Jan. 2012), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

[174] Johannes Kögel, Jennifer R Schmid, Ralf J Jox, and Orsolya Friedrich. 2019. Using brain-computer interfaces: a scoping review of studies employing social research methods. *BMC medical ethics* 20, 1 (2019), 1–17.

[175] Nataliya Kosmyna and Anatole Lécuyer. 2019. A conceptual space for EEG-based brain-computer interfaces. *PloS one* 14, 1 (2019), e0210145.

[176] Nataliya Kosmyna and Anatole Lécuyer. 2017. Designing Guiding Systems for Brain-Computer Interfaces. *Frontiers in Human Neuroscience* 11 (July 2017). https://doi.org/10.3389/fnhum.2017.00396

[177] Nataliya Kosmyna and Anatole Lécuyer. 2019. A conceptual space for EEG-based brain-computer interfaces. *PLOS ONE* 14, 1 (Jan. 2019), e0210145. https://doi.org/10.1371/journal.pone.0210145

[178] Nataliya Kosmyna, Nathalie Soetaert, and Cassandra Scheirer. 2022. A Pilot Study of Using Brain-Computer Interfaces in Classrooms for Promoting Formal Educational Activities. In *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 3*, Kohei Arai (Ed.). Springer International Publishing, Cham, 590–609. https://doi.org/10.1007/978-3-030-89912-7_46

[179] Nataliya Kosmyna, Franck Tarpin-Bernard, and Bertrand Rivet. 2015. Adding Human Learning in Brain–Computer Interfaces (BCIs): Towards a Practical Control Modality. *ACM Trans. Comput.-Hum. Interact.* 22, 3, Article 12 (May 2015), 37 pages. https://doi.org/10.1145/2723162

[180] Christian Kothe, Seyed Yahya Shirazi, Tristan Stenner, David Medine, Chadwick Boulay, Matthew I. Grivich, Tim Mullen, Arnaud Delorme, and Scott Makeig. 2024. The Lab Streaming Layer for Synchronized Multimodal Recording. https://doi.org/10.1101/2024.02.13.580071

[181] Christian Andreas Kothe and Scott Makeig. 2013. BCILAB: a platform for brain–computer interface development. *Journal of Neural Engineering* 10, 5 (Aug. 2013), 056014. https://doi.org/10.1088/1741-2560/10/5/056014

[182] Steve W. J. Kozlowski, Georgia T. Chao, James A. Grand, Michael T. Braun, and Goran Kuljanin. 2013. Advancing Multilevel Research Design: Capturing the Dynamics of Emergence. *Organizational Research Methods* 16, 4 (Oct. 2013), 581–615. https://doi.org/10.1177/1094428113493119

[183] Alexander Kraskov. 2004. Estimating mutual information. *Physical Review E* 69, 6 (2004). https://doi.org/10.1103/PhysRevE.69.066138

[184] Andrey Krekhov, Katharina Emmerich, Ronja Rotthaler, and Jens Krueger. 2021. Puzzles Unpuzzled: Towards a Unified Taxonomy for Analog and Digital Escape Room Games. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (Oct. 2021), 269:1–269:24. https://doi.org/10.1145/3474696

[185] Olave E Krigolson, Chad C Williams, Angela Norton, Cameron D Hassall, and Francisco L Colino. 2017. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in neuroscience* 11 (2017), 243179.

[186] Laurens R. Krol, Sarah-Christin Freytag, and Thorsten O. Zander. 2017. Meyendtris: a hands-free, multimodal tetris clone using eye tracking and passive BCI for intuitive neuroadaptive gaming. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 433–437. https://doi.org/10.1145/3136755.3136805

[187] Laurens R. Krol and Thorsten O. Zander. 2017. Passive BCI-based neuroadaptive systems. In *Proceedings of the 7th Graz Brain-Computer Interface Conference 2017*. Graz University of Technology. https://doi.org/10.3217/978-3-85125-533-1-46

[188] Dean J Krusienski, Eric W Sellers, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. 2008. Toward enhanced P300 speller performance. *Journal of neuroscience methods* 167, 1 (2008), 15–21.

[189] Nitin Kumar, Venkata Nishanth Lolla, Eamonn Keogh, Stefano Lonardi, Chotirat Ann Ratanamahatana, and Li Wei. 2005. Time-series bitmaps: a practical visualization tool for working with large time series databases. In *Proceedings of the 2005 SIAM international conference on data mining*. SIAM, 531–535.

[190] Shitij Kumar and Ferat Sahin. 2015. A framework for a real time intelligent and interactive Brain Computer Interface. *Computers & Electrical Engineering* 43 (April 2015), 193–214. https://doi.org/10.1016/j.compeleceng.2015.03.024

[191] Terri R. Kurtzberg and Teresa M. Amabile. 2001. From Guilford to Creative Synergy: Opening the Black Box of Team-Level Creativity. *Creativity Research Journal* 13, 3-4 (Oct. 2001), 285–294. https://doi.org/10.1207/S15326934CRJ1334_06

[192] Thomas Lampe, Lukas D.J. Fiederer, Martin Voelker, Alexander Knorr, Martin Riedmiller, and Tonio Ball. 2014. A brain-computer interface for high-level remote control of an autonomous, reinforcement-learning-based robotic system for reaching and

grasping. In *Proceedings of the 19th international conference on Intelligent User Interfaces (IUI '14)*. Association for Computing Machinery, New York, NY, USA, 83–88. https://doi.org/10.1145/2557500.2557533

[193] Johnny Chung Lee and Desney S. Tan. 2006. Using a Low-cost Electroencephalograph for Task Classification in HCI Research. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux, Switzerland) *(UIST '06)*. ACM, New York, NY, USA, 81–90. https://doi.org/10.1145/1166253.1166268

[194] Yi-Chieh Lee, Wen-Chieh Lin, Jung-Tai King, Li-Wei Ko, Yu-Ting Huang, and Fu-Yin Cherng. 2014. An EEG-based approach for evaluating audio notifications under ambient sounds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3817–3826. https://doi.org/10.1145/2556288.2557076

[195] Jaime Ranulfo Leite Filho, Thiago Adriano Coleti, and Marcelo Morandini. 2024. Evaluating chatbot user experience (UX) through electroencephalography measures: A systematic literature review. *Computers in Human Behavior Reports* 13 (March 2024), 100381. https://doi.org/10.1016/j.chbr.2024.100381

[196] G. Lightbody, M. Ware, P. McCullagh, M.D. Mulvenna, E. Thomson, S. Martin, D. Todd, V. C. Medina, and S. C. Martinez. 2010. A user centred approach for developing Brain-Computer Interfaces. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 1–8. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2010.8888 ISSN: 2153-1641.

[197] Chin-Teng Lin, Chun-Hsiang Chuang, Chih-Sheng Huang, Shu-Fang Tsai, Shao-Wei Lu, Yen-Hsuan Chen, and Li-Wei Ko. 2014. Wireless and Wearable EEG System for Evaluating Driver Vigilance. *IEEE Transactions on Biomedical Circuits and Systems* 8, 2 (April 2014), 165–176. https://doi.org/10.1109/TBCAS.2014.2316224

[198] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. 2003. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. 2–11.

[199] Jussi T. Lindgren, Adrien Merlini, Anatole Lécuyer, and Francesco P. Andriulli. 2018. simBCI—A Framework for Studying BCI Methods by Simulated EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 11 (Nov. 2018), 2096–2105. https://doi.org/10.1109/TNSRE.2018.2873061

[200] Fannie Liu, Laura Dabbish, and Geoff Kaufman. 2017. Can Biosignals be Expressive? How Visualizations Affect Impression Formation from Shared Brain Activity. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 71:1–71:21. https://doi.org/10.1145/3134706

[201] Ning Liu, Charis Mok, Emily E. Witt, Anjali H. Pradhan, Jingyuan E. Chen, and Allan L. Reiss. 2016. NIRS-Based Hyperscanning Reveals Inter-brain Neural Synchronization during Cooperative Jenga Game with Face-to-Face Communication. *Frontiers in Human Neuroscience* 10 (2016). https://www.frontiersin.org/articles/10.3389/fnhum.2016.00082

[202] Ruixue Liu, Erin Walker, Leah Friedman, Catherine M. Arrington, and Erin T. Solovey. 2021. fNIRS-based classification of mind-wandering with personalized window selection for multimodal learning interfaces. *Journal on Multimodal User Interfaces* 15, 3 (Sept. 2021), 257–272. https://doi.org/10.1007/s12193-020-00325-z

[203] Efstratios Livanis, Polychronis Voultsos, Konstantinos Vadikolias, Panagiotis Pantazakos, and Alexandra Tsaroucha. 2024. Understanding the Ethical Issues of Brain-Computer Interfaces (BCIs): A Blessing or the Beginning of a Dystopian Future? *Cureus* 16, 4 (2024), e58243. https://doi.org/10.7759/cureus.58243

[204] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. 2018. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering* 15, 3 (April 2018), 031005. https://doi.org/10.1088/1741-2552/aab2f2

[205] Fabien Lotte, Josef Faller, Christoph Guger, Yann Renard, Gert Pfurtscheller, Anatole Lécuyer, and Robert Leeb. 2013. Combining BCI with Virtual Reality: Towards New Applications and Improved BCI. In *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications*, Brendan Z. Allison, Stephen Dunne, Robert Leeb, José Del R. Millán, and Anton Nijholt (Eds.). Springer, Berlin, Heidelberg, 197–220. https://doi.org/10.1007/978-3-642-29746-5_10

[206] Kelong Lu and Ning Hao. 2019. When do we fall in neural synchrony with others? *Social Cognitive and Affective Neuroscience* 14, 3 (March 2019), 253–261. https://doi.org/10.1093/scan/nsz012

[207] Kelong Lu, Xinuo Qiao, and Ning Hao. 2019. Praising or keeping silent on partner's ideas: Leading brainstorming in particular ways. *Neuropsychologia* 124 (Feb. 2019), 19–30. https://doi.org/10.1016/j.neuropsychologia.2019.01.004

[208] Michael Lührs and Rainer Goebel. 2017. Turbo-Satori: a neurofeedback and brain–computer interface toolbox for real-time functional near-infrared spectroscopy. *Neurophotonics* 4, 4 (2017), 041504–041504.

[209] Kristiyan Lukanov, Horia A. Maior, and Max L. Wilson. 2016. Using fNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 4011–4016. https://doi.org/10.1145/2858036.2858236

[210] Paul Léné, Alexander J. Karran, Elise Labonté-Lemoyne, Sylvain Sénécal, Marc Fredette, Kevin J. Johnson, and Pierre-Majorique Léger. 2021. Is there collaboration

specific neurophysiological activation during collaborative task activity? An analysis of brain responses using electroencephalography and hyperscanning. *Brain and Behavior* 11, 11 (2021), e2270. https://doi.org/10.1002/brb3.2270

[211] Sergio López Bernal, Juan Antonio Martínez López, Enrique Tomás Martínez Beltrán, Mario Quiles Pérez, Gregorio Martínez Pérez, and Alberto Huertas Celdrán. 2024. NeuronLab: BCI framework for the study of biosignals. *Neurocomputing* (June 2024), 128027. https://doi.org/10.1016/j.neucom.2024.128027

[212] Markus Löning, Anthony Bagnall, Sajaysurya Ganesh, Viktor Kazakov, Jason Lines, and Franz J. Király. 2019. sktime: A Unified Interface for Machine Learning with Time Series. https://arxiv.org/abs/1909.07872v1

[213] Xinyao Ma, Zhaolin Yao, Yijun Wang, Weihua Pei, and Hongda Chen. 2018. Combining Brain-Computer Interface and Eye Tracking for High-Speed Text Entry in Virtual Reality. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. ACM, New York, NY, USA, 263–267. https://doi.org/10.1145/3172944.3172988

[214] Horia A Maior, Richard Ramchurn, Sarah Martindale, Ming Cai, Max L Wilson, and Steve Benford. 2019. fNIRS and Neurocinematics. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[215] Horia A. Maior, Max L. Wilson, and Sarah Sharples. 2018. Workload Alerts—Using Physiological Measures of Mental Workload to Provide Feedback During Tasks. *ACM Trans. Comput.-Hum. Interact.* 25, 2, Article 9 (apr 2018), 30 pages. https://doi.org/10.1145/3173380

[216] AI Makarova and VV Sulimova. 2017. High-performance DTW-based signals comparison for the brain electroencephalograms analysis. In *CEUR Workshop Proceedings*. 18–24.

[217] Christopher J Markiewicz, Krzysztof J Gorgolewski, Franklin Feingold, Ross Blair, Yaroslav O Halchenko, Eric Miller, Nell Hardcastle, Joe Wexler, Oscar Esteban, Mathias Goncavles, Anita Jwa, and Russell Poldrack. 2021. The OpenNeuro resource for sharing of neuroscience data. *eLife* 10 (oct 2021), e71774. https://doi.org/10.7554/eLife.71774

[218] Michelle A. Marks, John E. Mathieu, and Stephen J. Zaccaro. 2001. A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review* 26, 3 (2001), 356–376. https://doi.org/10.2307/259182

[219] Wendy Martinez, Johann Benerradi, Serena Midha, Horia A. Maior, and Max L. Wilson. 2022. Understanding the Ethical Concerns for Neurotechnology in the Future of Work. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3533406.3533423

[220] S.G. Mason and G.E. Birch. 2003. A general framework for brain-computer interface design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11, 1 (March 2003), 70–85. https://doi.org/10.1109/TNSRE.2003.810426

[221] John E. Mathieu, Mikhail A. Wolfson, Semin Park, Margaret M. Luciano, Wendy L. Bedwell-Torres, P. Scott Ramsay, Elizabeth A. Klock, and Scott I. Tannenbaum. 2022. Indexing dynamic collective constructs using computer-aided text analysis: Construct validity evidence and illustrations featuring team processes. *Journal of Applied Psychology* 107, 4 (2022), 533–559. https://doi.org/10.1037/apl0000856

[222] Fiachra Matthews, Christopher Soraghan, Tomas E. Ward, Charles Markham, and Barak A. Pearlmutter. 2008. Software platform for rapid prototyping of NIRS brain computer interfacing techniques. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 4840–4843. https://doi.org/10.1109/IEMBS.2008.4650297 ISSN: 1558-4615.

[223] Roger C. Mayer and Mark B. Gavin. 2005. Trust in Management and Performance: Who Minds the Shop While the Employees Watch the Boss? *Academy of Management Journal* 48, 5 (Oct. 2005), 874–888. https://doi.org/10.5465/amj.2005.18803928

[224] Oded Mayo and Ilanit Gordon. 2020. In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology* 57, 6 (2020), e13574. https://doi.org/10.1111/psyp.13574

[225] Daniel J. McAllister. 1995. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *The Academy of Management Journal* 38, 1 (1995), 24–59. https://doi.org/10.2307/256727

[226] Janetta Mitchell McCoy and Gary W. Evans. 2002. The Potential Role of the Physical Environment in Fostering Creativity. *Creativity Research Journal* 14, 3-4 (Oct. 2002), 409–426. https://doi.org/10.1207/S15326934CRJ1434_11

[227] Kenneth R. McLeroy, Whitney Garney, Evan Mayo-Wilson, and Sean Grant. 2016. Scientific Reporting: Raising the Standards. *Health Education & Behavior* 43, 5 (Oct. 2016), 501–508. https://doi.org/10.1177/1090198116668522

[228] Timothy McMahan, Ian Parberry, and Thomas D. Parsons. 2015. Evaluating Player Task Engagement and Arousal Using Electroencephalography. *Procedia Manufacturing* 3 (2015), 2303–2310. https://doi.org/10.1016/j.promfg.2015.07.376 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

[229] Meert, Wannes, Hendrickx, Killian, Aras Yurtman, Pieter Robberechts, Dany Vohl, Eric Ma, Gust Verbruggen, Marco Rossi, Mazhar Shaikh, Muhammad Yasirroni, Todd, Wojciech Zieliński, Toon Van Craenendonck, and Sai Wu. 2022. wannesm/dtaidistance: v2.3.5. https://zenodo.org/records/5901139

[230] Nick Merrill and John Chuang. 2018. From Scanning Brains to Reading Minds: Talking to Engineers about Brain-Computer Interface. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3173574.3173897

[231] Serena Midha, Horia A. Maior, Max L. Wilson, and Sarah Sharples. 2021. Measuring Mental Workload Variations in Office Work Tasks using fNIRS. *International Journal of Human-Computer Studies* 147 (2021), 102580. https://doi.org/10.1016/j.ijhcs.2020.102580

[232] Vojkan Mihajlović. 2019. EEG spectra vs recurrence features in understanding cognitive effort. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers (ISWC '19)*. Association for Computing Machinery, New York, NY, USA, 160–165. https://doi.org/10.1145/3341163.3347746

[233] Michael P Milham, R Cameron Craddock, Jake J Son, Michael Fleischmann, Jon Clucas, Helen Xu, Bonhwang Koo, Anirudh Krishnakumar, Bharat B Biswal, F Xavier Castellanos, et al. 2018. Assessment of the impact of shared brain imaging data on the scientific literature. *Nature communications* 9 (2018).

[234] Felip Miralles, Eloisa Vargiu, Xavier Rafael-Palou, Marc Solà, Stefan Dauwalder, Christoph Guger, Christoph Hintermüller, Arnau Espinosa, Hannah Lowish, Suzanne Martin, Elaine Armstrong, and Jean Daly. 2015. Brain–Computer Interfaces on Track to Home: Results of the Evaluation at Disabled End-Users' Homes and Lessons Learnt. *Frontiers in ICT* 2 (Nov. 2015). https://doi.org/10.3389/fict.2015.00025

[235] Evan Morgan, Hatice Gunes, and Nick Bryan-Kinns. 2015. Using affective and behavioural sensors to explore aspects of collaborative music making. *International Journal of Human-Computer Studies* 82 (Oct. 2015), 31–47. https://doi.org/10.1016/j.ijhcs.2015.05.002

[236] Ekaterina Muhl and Roberto Andorno. 2023. Neurosurveillance in the workplace: do employers have the right to monitor employees' minds? *Frontiers in Human Dynamics* 5 (Sept. 2023). https://doi.org/10.3389/fhumd.2023.1245619

[237] Tamara Munzner. 2014. *Visualization Analysis and Design*. CRC press.

[238] Andrew Myrden and Tom Chau. 2017. A passive EEG-BCI for single-trial detection of changes in mental state. *IEEE Transactions on neural systems and rehabilitation Engineering* 25, 4 (2017), 345–356.

[239] Gattigorla Nagendar and CV Jawahar. 2015. Efficient word image retrieval using fast DTW distance. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 876–880.

[240] Rodica Neamtu, Ramoza Ahsan, Charles Lovering, Cuong Nguyen, Elke Rundensteiner, and Gabor Sarkozy. 2017. Interactive time series analytics powered by ONEX.

In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1595–1598.

[241] Rodica Neamtu, Ramoza Ahsan, Cuong Dinh Tri Nguyen, Charles Lovering, Elke A Rundensteiner, and Gabor Sarkozy. 2020. A General Approach For Supporting Time Series Matching using Multiple-Warped Distances. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[242] Rodica Neamtu, Ramoza Ahsan, Elke Rundensteiner, and Gabor Sarkozy. 2016. Interactive time series exploration powered by the marriage of similarity distances. *Proceedings of the VLDB Endowment* 10, 3 (2016), 169–180.

[243] Rodica Neamtu, Ramoza Ahsan, Elke A Rundensteiner, Gabor Sarkozy, Eamonn Keogh, Hoang Anh Dau, Cuong Nguyen, and Charles Lovering. 2018. Generalized dynamic time warping: Unleashing the warping power hidden in point-wise distances. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 521–532.

[244] Cuong Nguyen, Charles Lovering, and Rodica Neamtu. 2017. Ranked time series matching by interleaving similarity distances. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 3530–3539.

[245] Trinh Nguyen, Hanna Schleihauf, Ezgi Kayhan, Daniel Matthes, Pascal Vrtička, and Stefanie Hoehl. 2020. The effects of interaction quality on neural synchrony during mother-child problem solving. *Cortex* 124 (March 2020), 235–249. https://doi.org/10.1016/j.cortex.2019.11.020

[246] Trinh Nguyen, Hanna Schleihauf, Melanie Kungl, Ezgi Kayhan, Stefanie Hoehl, and Pascal Vrtička. 2021. Interpersonal Neural Synchrony During Father–Child Problem Solving: An fNIRS Hyperscanning Study. *Child Development* 92, 4 (2021), e565–e580. https://doi.org/10.1111/cdev.13510

[247] Femke Nijboer, Danny Plass-Oude Bos, Yvonne Blokland, René van Wijk, and Jason Farquhar. 2014. Design requirements and potential target users for brain-computer interfaces – recommendations from rehabilitation professionals. *Brain-Computer Interfaces* (Jan. 2014). https://www.tandfonline.com/doi/abs/10.1080/2326263X.2013.877210

[248] Bernard A. Nijstad and Wolfgang Stroebe. 2006. How the Group Affects the Mind: A Cognitive Model of Idea Generation in Groups. *Personality and Social Psychology Review* 10, 3 (Aug. 2006), 186–213. https://doi.org/10.1207/s15327957pspr1003_1

[249] Takayuki Nozawa, Yukako Sasaki, Kohei Sakaki, Ryoichi Yokoyama, and Ryuta Kawashima. 2016. Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *NeuroImage* 133 (June 2016), 484–497. https://doi.org/10.1016/j.neuroimage.2016.03.059

[250] NSF - National Science Foundation. [n. d.]. Broader Impacts. https://www.nsf.gov/funding/learn/broader-impacts

[251] Joseph K. Nuamah, Younho Seong, and Sun Yi. 2017. Electroencephalography (EEG) classification of cognitive tasks based on task engagement index. In *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. 1–6. https://doi.org/10.1109/COGSIMA.2017.7929581

[252] Marc R. Nuwer, Giancarlo Comi, Ronald Emerson, Anders Fuglsang-Frederiksen, Jean-Michel Guérit, Hermann Hinrichs, Akio Ikeda, Fransisco Jose C. Luccas, and Peter Rappelsburger. 1998. IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology* 106, 3 (March 1998), 259–261. https://doi.org/10.1016/S0013-4694(97)00106-5

[253] Pablo Núñez, Jesús Poza, Carlos Gómez, Verónica Barroso-García, Saúl J. Ruiz-Gómez, Aarón Maturana-Candelas, Miguel A. Tola-Arribas, Mónica Cano, and Roberto Hornero. 2019. Characterization of EEG Resting-state Activity in Alzheimer's Disease by Means of Recurrence Plot Analyses. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5786–5789. https://doi.org/10.1109/EMBC.2019.8856600 ISSN: 1558-4615.

[254] Qasem Obeidat, Tom Campbell, and Jun Kong. 2013. The Zigzag Paradigm: A New P300-based Brain Computer Interface. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (Sydney, Australia) *(ICMI '13)*. ACM, New York, NY, USA, 205–212. https://doi.org/10.1145/2522848.2522880

[255] Alejandro Ojeda, Nathalie Buscher, Pragathi Balasubramani, Vojislav Maric, Dhakshin Ramanathan, and Jyoti Mishra. 2020. SimBSI: An open-source Simulink library for developing closed-loop brain signal interfaces in animals and humans. *Biomedical Physics & Engineering Express* 6, 3 (April 2020), 035023. https://doi.org/10.1088/2057-1976/ab6e20

[256] Sally Olderbak, Andero Uusberg, Carolyn MacCann, Katja M. Pollak, and James J. Gross. 2022. The Process Model of Emotion Regulation Questionnaire: Assessing Individual Differences in Strategy Stage and Orientation. *Assessment* (Dec. 2022), 10731911221134601. https://doi.org/10.1177/10731911221134601

[257] Giuseppe Orlando and Giovanna Zimatore. 2018. Recurrence quantification analysis of business cycles. *Chaos, Solitons & Fractals* 110 (May 2018), 82–94. https://doi.org/10.1016/j.chaos.2018.02.032

[258] Paul Paulus. 2000. Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. *Applied Psychology* 49, 2 (2000), 237–262. https://doi.org/10.1111/1464-0597.00013

[259] Paul B. Paulus and Mary Dzindolet. 2008. Social influence, creativity and innovation. *Social Influence* 3, 4 (Dec. 2008), 228–247. https://doi.org/10.1080/15534510802341082

[260] Paul B. Paulus, Mary Dzindolet, and Nicholas W. Kohn. 2012. Chapter 14 - Collaborative Creativity—Group Creativity and Team Innovation. In *Handbook of Organizational Creativity*, Michael D. Mumford (Ed.). Academic Press, San Diego, 327–357. https://doi.org/10.1016/B978-0-12-374714-3.00014-8

[261] Paul B. Paulus and Jared B. Kenworthy. 2021. Research Findings on Ideational Creativity in Groups. In *Creativity and Innovation: Cognitive, Social, and Computational Approaches*, Simona Doboli, Jared B. Kenworthy, Ali A. Minai, and Paul B. Paulus (Eds.). Springer International Publishing, Cham, 47–67. https://doi.org/10.1007/978-3-030-77198-0_3

[262] Evan M M. Peck, Beste F. Yuksel, Alvitta Ottley, Robert J.K. Jacob, and Remco Chang. 2013. Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 473–482. https://doi.org/10.1145/2470654.2470723

[263] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[264] Michelle Pham, Sara Goering, Matthew Sample, Jane E. Huggins, and Eran Klein. 2018. Asilomar survey: researcher perspectives on ethical principles and guidelines for BCI research. *Brain-Computer Interfaces* (Oct. 2018). https://www.tandfonline.com/doi/abs/10.1080/2326263X.2018.1530010

[265] Aaron A Phillips, Franco HN Chan, Mei Mu Zi Zheng, Andrei V Krassioukov, and Philip N Ainslie. 2016. Neurovascular coupling in humans: Physiology, methodological advances and clinical implications. *Journal of Cerebral Blood Flow & Metabolism* 36, 4 (2016), 647–664. https://doi.org/10.1177/0271678X15617954 arXiv:https://doi.org/10.1177/0271678X15617954 PMID: 26661243.

[266] Matthew Pike, Richard Ramchurn, Steve Benford, and Max L. Wilson. 2016. #Scanners: Exploring the Control of Adaptive Films using Brain-Computer Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5385–5396. https://doi.org/10.1145/2858036.2858276

[267] Ming-Zher Poh, Daniel McDuff, and Rosalind Picard. 2011. A medical mirror for non-contact health monitoring. In *ACM SIGGRAPH 2011 Emerging Technologies (SIGGRAPH '11)*. Association for Computing Machinery, New York, NY, USA, 1. https://doi.org/10.1145/2048259.2048261

[268] Russell A Poldrack, Aniket Kittur, Donald Kalar, Eric Miller, Christian Seppa, Yolanda Gil, D Stott Parker, Fred W Sabb, and Robert M Bilder. 2011. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in neuroinformatics* 5 (2011), 17.

[269] Kathrin Pollmann, Daniel Ziegler, Matthias Peissner, and Mathias Vukelić. 2017. A New Experimental Paradigm for Affective Research in Neuro-Adaptive Technologies. In *Proceedings of the 2017 ACM Workshop on An Application-Oriented Approach to BCI out of the Laboratory* (Limassol, Cyprus) *(BCIforReal '17)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3038439.3038442

[270] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. 1995. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology* 40, 1 (May 1995), 187–195. https://doi.org/10.1016/0301-0511(95)05116-3

[271] Emily Postan. 2021. Narrative Devices: Neurotechnologies, Information, and Self-Constitution. *Neuroethics* 14, 2 (July 2021), 231–251. https://doi.org/10.1007/s12152-020-09449-1

[272] Mirjana Prpa and Philippe Pasquier. 2019. Brain-Computer Interfaces in Contemporary Art: A State of the Art and Taxonomy. In *Brain Art: Brain-Computer Interfaces for Artistic Expression*, Anton Nijholt (Ed.). Springer International Publishing, Cham, 65–115. https://doi.org/10.1007/978-3-030-14323-7_3

[273] Felix Putze, Christoph Amma, and Tanja Schultz. 2015. Design and Evaluation of a Self-Correcting Gesture Interface Based on Error Potentials from EEG. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 3375–3384. https://doi.org/10.1145/2702123.2702184

[274] Felix Putze, Tilman Ihrig, Tanja Schultz, and Wolfgang Stuerzlinger. 2020. Platform for Studying Self-Repairing Auto-Corrections in Mobile Text Entry based on Brain Activity, Gaze, and Context. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376815

[275] Felix Putze, Susanne Putze, Merle Sagehorn, Christopher Micek, and Erin T. Solovey. 2022. Understanding HCI Practices and Challenges of Experiment Reporting with Brain Signals: Towards Reproducibility and Reuse. *ACM Transactions on Computer-Human Interaction* 29, 4 (March 2022), 31:1–31:43. https://doi.org/10.1145/3490554

[276] Felix Putze, Susanne Putze, Merle Sagehorn, Christopher Micek, and Erin T Solovey. 2022. Understanding hci practices and challenges of experiment reporting with brain signals: Towards reproducibility and reuse. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–43.

[277] Felix Putze and Tanja Schultz. 2014. Investigating Intrusiveness of Workload Adaptation. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. Association for Computing Machinery, New York, NY, USA, 275–281. https://doi.org/10.1145/2663204.2663279

[278] Felix Putze, Maik Schünemann, Tanja Schultz, and Wolfgang Stuerzlinger. 2017. Automatic Classification of Auto-correction Errors in Predictive Text Entry Based on EEG and Context Information. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. ACM, New York, NY, USA, 137–145. https://doi.org/10.1145/3136755.3136784

[279] Susanne Putze, Robert Porzel, Gian-Luca Savino, and Rainer Malaka. 2018. A Manageable Model for Experimental Research Data: An Empirical Study in the Materials Sciences. In *Advanced Information Systems Engineering (Lecture Notes in Computer Science)*, John Krogstie and Hajo A. Reijers (Eds.). Springer International Publishing, 424–439.

[280] Melissa Quek, Daniel Boland, John Williamson, Roderick Murray-Smith, Michele Tavella, Serafeim Perdikis, Martijn Schreuder, and Michael Tangermann. 2011. Simulating the feel of brain-computer interfaces for design, development and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 25–28. https://doi.org/10.1145/1978942.1978947

[281] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 262–270.

[282] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2013. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 7, 3 (2013), 1–31.

[283] Richard Ramchurn, Sarah Martindale, Max L. Wilson, and Steve Benford. 2019. From Director's Cut to User's Cut: to Watch a Brain-Controlled Film is to Edit it. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300378

[284] Diego A Reinero, Suzanne Dikker, and Jay J Van Bavel. 2021. Inter-brain synchrony in teams predicts collective performance. *Social cognitive and affective neuroscience* 16, 1-2 (2021), 43–57.

[285] Yann Renard, Fabien Lotte, Guillaume Gibert, Marco Congedo, Emmanuel Maby, Vincent Delannoy, Olivier Bertrand, and Anatole Lécuyer. 2010. OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. *Presence* 19, 1 (Feb. 2010), 35–53. https://doi.org/10.1162/pres.19.1.35

[286] Daniel C. Richardson and Rick Dale. 2005. Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science* 29, 6 (2005), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

[287] Christoph Riedl, Young Ji Kim, Pranav Gupta, Thomas W. Malone, and Anita Williams Woolley. 2021. Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences* 118, 21 (May 2021), e2005737118. https://doi.org/10.1073/pnas.2005737118

[288] Mathieu Rodrigue, Jungah Son, Barry Giesbrecht, Matthew Turk, and Tobias Höllerer. 2015. Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. ACM, New York, NY, USA, 121–125. https://doi.org/10.1145/2678025.2701382

[289] Michael Rosen. 2010. Collaborative Problem Solving: The Role Of Team Knowledge Building Processes And External Representations. *Electronic Theses and Dissertations* (Jan. 2010). https://stars.library.ucf.edu/etd/4213

[290] Kathy Ryall, Neal Lesh, Tom Lanning, Darren Leigh, Hiroaki Miyashita, and Shigeru Makino. 2005. QueryLines: Approximate Query for Visual Browsing. *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (2005).

[291] Matthew Sample, Sebastian Sattler, Stefanie Blain-Moraes, David Rodríguez-Arias, and Eric Racine. 2020. Do Publics Share Experts' Concerns about Brain–Computer Interfaces? A Trinational Survey on the Ethics of Neural Technology. *Science, Technology, & Human Values* 45, 6 (Nov. 2020), 1242–1270. https://doi.org/10.1177/0162243919879220

[292] Hendrik Santosa, Xuetong Zhai, Frank Fishburn, and Theodore Huppert. 2018. The NIRS brain AnalyzIR toolbox. *Algorithms* 11, 5 (2018), 73.

[293] Sebastian Sattler and Dana Pietralla. 2022. Public attitudes towards neurotechnology: Findings from two experiments concerning Brain Stimulation Devices (BSDs) and Brain-Computer Interfaces (BCIs). *PLOS ONE* 17, 11 (Nov. 2022), e0275454. https://doi.org/10.1371/journal.pone.0275454

[294] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. 2004. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering* 51, 6 (June 2004), 1034–1043. https://doi.org/10.1109/TBME.2004.827072

[295] Christina Schneegass, Thomas Kosch, Andrea Baumann, Marius Rusu, Mariam Hassib, and Heinrich Hussmann. 2020. BrainCoDe: Electroencephalography-based Comprehension Detection during Reading and Listening. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376707

[296] Thomas Schreiber and Andreas Schmitz. 2000. Surrogate time series. *Physica D: Nonlinear Phenomena* 142, 3 (Aug. 2000), 346–382. https://doi.org/10.1016/S0167-2789(00)00043-9

[297] Martijn Schreuder, Angela Riccio, Monica Risetti, Sven Dähne, Andrew Ramsay, John Williamson, Donatella Mattia, and Michael Tangermann. 2013. User-centered design in brain–computer interfaces—A case study. *Artificial Intelligence in Medicine* 59, 2 (Oct. 2013), 71–80. https://doi.org/10.1016/j.artmed.2013.07.005

[298] Nathan Semertzidis, Michaela Vranic-Peters, Josh Andres, Brahmi Dwivedi, Yutika Chandrashekhar Kulwe, Fabio Zambetta, and Florian Floyd Mueller. 2020. Neo-Noumena: Augmenting Emotion Communication. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376599

[299] Nathan Semertzidis, Fabio Zambetta, and Florian "Floyd" Mueller. 2023. Brain-Computer Integration: A Framework for the Design of Brain-Computer Interfaces from an Integrations Perspective. *ACM Transactions on Computer-Human Interaction* 30, 6 (Sept. 2023), 86:1–86:48. https://doi.org/10.1145/3603621

[300] Pavel Senin. 2008. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA* 855, 1-23 (2008), 40.

[301] Sonia Jawaid Shaikh and Ignacio F. Cruz. 2023. AI in human teams: effects on technology use, members' interactions, and creative performance under time scarcity. *AI & SOCIETY* 38, 4 (Aug. 2023), 1587–1600. https://doi.org/10.1007/s00146-021-01335-5

[302] Jaeyoung Shin, Alexander von Lühmann, Do-Won Kim, Jan Mehnert, Han-Jeong Hwang, and Klaus-Robert Müller. 2018. Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data* 5, 1 (Feb. 2018), 180003. https://doi.org/10.1038/sdata.2018.3

[303] Pao Siangliulue, Joel Chan, Steven P. Dow, and Krzysztof Z. Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 609–624. https://doi.org/10.1145/2984511.2984578

[304] Harshinder Singh, Misra , Neeraj, Hnizdo , Vladimir, Fedorowicz , Adam, , and Eugene Demchuk. 2003. Nearest Neighbor Estimates of Entropy. *American Journal of Mathematical and Management Sciences* 23, 3-4 (Feb. 2003), 301–321. https://doi.org/10.1080/01966324.2003.10737616

[305] Filip Škola and Fotis Liarokapis. 2019. Examining and Enhancing the Illusory Touch Perception in Virtual Reality Using Non-Invasive Brain Stimulation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 247:1–247:12. https://doi.org/10.1145/3290605.3300477

[306] Michael John Sebastian Smith. 1997. *Application-specific integrated circuits*. Vol. 7. Addison-Wesley Reading, MA.

[307] Erin Solovey, Paul Schermerhorn, Matthias Scheutz, Angelo Sassaroli, Sergio Fantini, and Robert Jacob. 2012. Brainput: Enhancing Interactive Systems with Streaming Fnirs Brain Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 2193–2202. https://doi.org/10.1145/2207676.2208372

[308] Erin Treacy Solovey, Daniel Afergan, Evan M. Peck, Samuel W. Hincks, and Robert J. K. Jacob. 2015. Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned Using fNIRS. *ACM Trans. Comput.-Hum. Interact.* 21, 6, Article 35 (Jan. 2015), 27 pages. https://doi.org/10.1145/2687926

[309] Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M. Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert J.K. Jacob. 2009. Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) *(UIST '09)*. ACM, New York, NY, USA, 157–166. https://doi.org/10.1145/1622176.1622207

[310] Erin T. Solovey and Felix Putze. 2021. Improving HCI with Brain Input: Review, Trends, and Outlook. *Foundations and Trends® in Human–Computer Interaction* 13, 4 (April 2021), 298–379. https://doi.org/10.1561/1100000078

[311] Martin Spüler, Tanja Krumpe, Carina Walter, Christian Scharinger, Wolfgang Rosenstiel, and Peter Gerjets. 2017. Brain-Computer Interfaces for Educational Applications. In *Informational Environments : Effects of Use, Effective Designs*, Jürgen Buder and Friedrich W. Hesse (Eds.). Springer International Publishing, Cham, 177–201. https://doi.org/10.1007/978-3-319-64274-1_8

[312] Pierce Stegman, Chris S. Crawford, Marvin Andujar, Anton Nijholt, and Juan E. Gilbert. 2020. Brain–Computer Interface Software: A Review and Discussion. *IEEE Transactions on Human-Machine Systems* 50, 2 (April 2020), 101–115. https://doi.org/10.1109/THMS.2020.2968411

[313] Robert J. Sternberg. 2006. Creating a vision of creativity: The first 25 years. *Psychology of Aesthetics, Creativity, and the Arts* S (2006), 2–12. https://doi.org/10.1037/1931-3896.S.1.2

[314] Robert J. Sternberg. 2018. A triangular theory of creativity. *Psychology of Aesthetics, Creativity, and the Arts* 12, 1 (2018), 50–67. https://doi.org/10.1037/aca0000095

[315] Ronald H. Stevens and Trysha L. Galloway. 2016. Modeling the neurodynamic organizations and interactions of teams. *Social Neuroscience* (March 2016). https://www.tandfonline.com/doi/abs/10.1080/17470919.2015.1056883

[316] Markus Andreas Stricker and Markus Orengo. 1995. Similarity of color images. In *Storage and retrieval for image and video databases III*, Vol. 2420. International Society for Optics and Photonics, 381–392.

[317] Xiao-yu Sun and Bin Ye. 2023. The functional differentiation of brain–computer interfaces (BCIs) and its ethical implications. *Humanities and Social Sciences Communications* 10, 1 (Nov. 2023), 1–9. https://doi.org/10.1057/s41599-023-02419-x

[318] Youqiang Sun, Jiuyong Li, Jixue Liu, Bingyu Sun, and Christopher Chow. 2014. An improvement of symbolic aggregate approximation distance measure for time series. *Neurocomputing* 138 (2014), 189–198.

[319] Stephanie Sutoko, Hiroki Sato, Atsushi Maki, Masashi Kiguchi, Yukiko Hirabayashi, Hirokazu Atsumori, Akiko Obata, Tsukasa Funane, and Takusige Katura. 2016. Tutorial on platform for optical topography analysis tools. *Neurophotonics* 3, 1 (January 2016), 010801. https://doi.org/10.1117/1.nph.3.1.010801

[320] Michael J Swain and Dana H Ballard. 1991. Color indexing. *International journal of computer vision* 7, 1 (1991), 11–32.

[321] Daniel Szafir and Bilge Mutlu. 2012. Pay Attention!: Designing Adaptive Agents That Monitor and Improve User Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/2207676.2207679

[322] Daniel Szafir and Bilge Mutlu. 2013. ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1001–1010. https://doi.org/10.1145/2470654.2466128

[323] Naoto Terasawa, Hiroki Tanaka, Sakriani Sakti, and Satoshi Nakamura. 2017. Tracking Liking State in Brain Activity While Watching Multiple Movies. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. ACM, New York, NY, USA, 321–325. https://doi.org/10.1145/3136755.3136772

[324] Thomas Terkildsen and Guido Makransky. 2019. Measuring presence in video games: An investigation of the potential use of physiological measures as indicators of presence. *International Journal of Human-Computer Studies* 126 (2019), 64 – 80. https://doi.org/10.1016/j.ijhcs.2019.02.006

[325] Amanda L. Thayer, Alexandra Petruzzelli, and Caitlin E. McClurg. 2018. Addressing the paradox of the team innovation process: A review and practical considerations. *American Psychologist* 73 (2018), 363–375. https://doi.org/10.1037/amp0000310

[326] Lauren C. Tindale, Derek Chiu, Nicole Minielly, Viorica Hrincu, Aline Talhouk, and Judy Illes. 2022. Wearable Biosensors in the Workplace: Perceptions and Perspectives. *Frontiers in Digital Health* 4 (2022). https://www.frontiersin.org/articles/10.3389/fdgth.2022.800367

[327] Keith F. Tipton, Richard N. Armstrong, Barbara M. Bakker, Amos Bairoch, Athel Cornish-Bowden, Peter J. Halling, Jan-Hendrik Hofmeyr, Thomas S. Leyh, Carsten Kettner, Frank M. Raushel, Johann Rohwer, Dietmar Schomburg, and Christoph Steinbeck. [n. d.]. Standards for Reporting Enzyme Data: The STRENDA Consortium: What It Aims to Do and Why It Should Be Helpful. 1, 1 ([n. d.]), 131–137. https://doi.org/10.1016/j.pisc.2014.02.012

[328] D. A. Todd, P. J. McCullagh, M. D. Mulvenna, and G. Lightbody. 2012. Investigating the use of brain-computer interaction to facilitate creativity. In *Proceedings of the 3rd Augmented Human International Conference (AH '12)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2160125.2160144

[329] Vinh Phuc Tran, Adel Ali Al-Jumaily, and Syed Mohammed Shamsul Islam. 2019. Doppler Radar-Based Non-Contact Health Monitoring for Obstructive Sleep Apnea Diagnosis: A Comprehensive Review. *Big Data and Cognitive Computing* 3, 1 (March 2019), 3. https://doi.org/10.3390/bdcc3010003

[330] Silvia Traunwieser. 2024. Neurodata-based headsets for the (digital) employee well-being – responsibilities between benefit and harm. *International Journal of Ethics and Systems* 41, 1 (Oct. 2024), 64–87. https://doi.org/10.1108/IJOES-02-2024-0044

[331] Matthias S. Treder and Benjamin Blankertz. 2010. (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions* 6, 1 (May 2010), 28. https://doi.org/10.1186/1744-9081-6-28

[332] Muhammad Umair, Corina Sas, and Miquel Alfaras. 2020. ThermoPixels: Toolkit for Personalizing Arousal-based Interfaces through Hybrid Crafting. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1017–1032. https://doi.org/10.1145/3357236.3395512

[333] Robin R. Vallacher, Andrzej Nowak, and Michal Zochowski. 2005. Dynamics of social coordination: The synchronization of internal states in close relationships. *Interaction*

Studies. Social Behaviour and Communication in Biological and Artificial Systems 6, 1 (April 2005), 35–52. https://doi.org/10.1075/is.6.1.04val

[334] Gabriel Alves Mendes Vasiljevic and Leonardo Cunha de Miranda. 2024. Model, Taxonomy and Methodology for Research Employing EEG-based Brain-Computer Interface Games. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems (IHC '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3638067.3638114

[335] Boris M. Velichkovsky and John Paulin Hansen. 1996. New Technological Windows into Mind: There is More in Eyes and Brains for Human-computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, British Columbia, Canada) *(CHI '96)*. ACM, New York, NY, USA, 496–503. https://doi.org/10.1145/238386.238619

[336] Chi Thanh Vi, Izdihar Jamil, David Coyle, and Sriram Subramanian. 2014. Error Related Negativity in Observing Interactive Tasks. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 3787–3796. https://doi.org/10.1145/2556288.2557015

[337] Ruben Vicente-Saez and Clara Martinez-Fuentes. 2018. Open Science now: A systematic literature review for an integrated definition. *Journal of business research* 88 (2018), 428–436.

[338] Carmen Vidaurre, Tilmann H. Sander, and Alois Schlögl. 2011. BioSig: The Free and Open Source Software Library for Biomedical Signal Processing. *Computational Intelligence and Neuroscience* 2011, 1 (2011), 935364. https://doi.org/10.1155/2011/935364

[339] Arno Villringer, J Planck, C Hock, L Schleinkofer, and U Dirnagl. 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience letters* 154, 1-2 (1993), 101–104.

[340] Michail Vlachos, Marios Hadjieleftheriou, Dimitrios Gunopulos, and Eamonn Keogh. 2003. Indexing multi-dimensional time-series with support for multiple distance measures. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 216–225.

[341] Rutger J. Vlek, David Steines, Dyana Szibbo, Andrea Kübler, Mary-Jane Schneider, Pim Haselager, and Femke Nijboer. 2012. Ethical Issues in Brain–Computer Interface Research, Development, and Dissemination. *Journal of Neurologic Physical Therapy* 36, 2 (June 2012), 94. https://doi.org/10.1097/NPT.0b013e31825064cc

[342] Emmanuelle Volle. 2018. Associative and Controlled Cognition in Divergent Thinking: Theoretical, Experimental, Neuroimaging Evidence, and New Directions. In *The Cambridge Handbook of the Neuroscience of Creativity*, Oshin Vartanian and

Rex E. Jung (Eds.). Cambridge University Press, Cambridge, 333–360. https://doi.org/10.1017/9781316556238.020

[343] Alexander von Lühmann, Xinge Li, Natalie Gilmore, David A. Boas, and Meryem A. Yücel. 2020. Open Access Multimodal fNIRS Resting State Dataset With and Without Synthetic Hemodynamic Responses. *Frontiers in Neuroscience* 14 (Sept. 2020). https://doi.org/10.3389/fnins.2020.579353

[344] Hana Vrzakova, Mary Jean Amon, Angela E. B. Stewart, and Sidney K. D'Mello. 2019. Dynamics of Visual Attention in Multiparty Collaborative Problem Solving using Multidimensional Recurrence Quantification Analysis. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300572

[345] Filip Škola and Fotis Liarokapis. 2019. Examining and Enhancing the Illusory Touch Perception in Virtual Reality Using Non-Invasive Brain Stimulation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 247, 12 pages. https://doi.org/10.1145/3290605.3300477

[346] Angela Vujic, Shreyas Nisal, and Pattie Maes. 2023. Joie: a Joy-based Brain-Computer Interface (BCI). In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3586183.3606761

[347] Dhaval Vyas, Dirk Heylen, Anton Nijholt, and Gerrit van der Veer. 2009. Collaborative Practices that Support Creativity in Design. In *ECSCW 2009*, Ina Wagner, Hilda Tellioğlu, Ellen Balka, Carla Simone, and Luigina Ciolfi (Eds.). Springer, London, 151–170. https://doi.org/10.1007/978-1-84882-854-4_9

[348] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[349] Ruth Wageman, Heidi Gardner, and Mark Mortensen. 2012. The changing ecology of teams: New directions for teams research. *Journal of Organizational Behavior* 33, 3 (2012), 301–315. https://doi.org/10.1002/job.1775

[350] Sebastian Wallot and Giuseppe Leonardi. 2018. Analyzing Multivariate Dynamics Using Cross-Recurrence Quantification Analysis (CRQA), Diagonal-Cross-Recurrence Profiles (DCRP), and Multidimensional Recurrence Quantification Analysis (MdRQA) – A Tutorial in R. *Frontiers in Psychology* 9 (2018). https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02232

[351] Sebastian Wallot and Dan Mønster. 2018. Calculation of Average Mutual Information (AMI) and False-Nearest Neighbors (FNN) for the Estimation of Embedding Parameters of Multidimensional Time Series in Matlab. *Frontiers in Psychology* 9 (Sept. 2018). https://doi.org/10.3389/fpsyg.2018.01679

[352] Sebastian Wallot, Andreas Roepstorff, and Dan Mønster. 2016. Multidimensional Recurrence Quantification Analysis (MdRQA) for the Analysis of Multidimensional Time-Series: A Software Implementation in MATLAB and Its Application to Group-Level Data in Joint Action. *Frontiers in Psychology* 7 (Nov. 2016). https://doi.org/10.3389/fpsyg.2016.01835

[353] André P. Walton. 2003. The impact of interpersonal factors on creativity. *International Journal of Entrepreneurial Behavior & Research* 9, 4 (Jan. 2003), 146–162. https://doi.org/10.1108/13552550310485120

[354] Kai Wang and Jeffrey V. Nickerson. 2017. A literature review on individual creativity support systems. *Computers in Human Behavior* 74 (Sept. 2017), 139–151. https://doi.org/10.1016/j.chb.2017.04.035

[355] Liang Wang, Zhe Huang, Ziyu Zhou, Devon McKeon, Giles Blaney, Michael C. Hughes, and Robert J. K. Jacob. 2021. *Taming FNIRS-Based BCI Input for Better Calibration and Broader Use.* Association for Computing Machinery, New York, NY, USA, 179–197. https://doi.org/10.1145/3472749.3474743

[356] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (April 2004), 600–612. https://doi.org/10.1109/TIP.2003.819861

[357] Andrew Warr and Eamonn O'Neill. 2006. The effect of group composition on divergent thinking in an interaction design activity. In *Proceedings of the 6th conference on Designing Interactive systems (DIS '06)*. Association for Computing Machinery, New York, NY, USA, 122–131. https://doi.org/10.1145/1142405.1142427

[358] Charles Webber and Joseph Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences* (Jan. 2005).

[359] Andreas Widmann, Erich Schröger, and Burkhard Maess. 2015. Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods* 250 (July 2015), 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

[360] John Williamson, Roderick Murray-Smith, Benjamin Blankertz, Matthias Krauledat, and K-R Müller. 2009. Designing for uncertain, asymmetric control: Interaction design for brain–computer interfaces. *International Journal of Human-Computer Studies* 67, 10 (Oct. 2009), 827–841. https://doi.org/10.1016/j.ijhcs.2009.05.009

[361] Jeanne M. Wilson, Thomas D. Fletcher, Tuck Pescosolido, and Debra A. Major. 2021. Extraversion and Leadership Emergence: Differences in Virtual and Face-to-Face Teams. *Small Group Research* 52, 5 (Oct. 2021), 535–564. https://doi.org/10.1177/1046496420986620

[362] Katherine Wolstencroft, Stuart Owen, Olga Krebs, Quyen Nguyen, Natalie J. Stanford, Martin Golebiewski, Andreas Weidemann, Meik Bittkowski, Lihua An, David Shockley,

Jacky L. Snoep, Wolfgang Mueller, and Carole Goble. 2015. SEEK: a systems biology data and model management platform. *BMC Systems Biology* 9, 1 (July 2015), 33. https://doi.org/10.1186/s12918-015-0174-y

[363] Krist Wongsuphasawat and Ben Shneiderman. 2009. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 27–34.

[364] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (Oct. 2010), 686–688. https://doi.org/10.1126/science.1193147

[365] Yong Xu, Harry L Graber, and Randall L Barbour. 2014. nirsLAB: a computing environment for fNIRS neuroimaging data analysis. In *Biomedical Optics*. Optica Publishing Group, BM3A–1.

[366] Hua Xue, Kelong Lu, and Ning Hao. 2018. Cooperation makes two less-creative individuals turn into a highly-creative pair. *NeuroImage* 172 (May 2018), 527–537. https://doi.org/10.1016/j.neuroimage.2018.02.007

[367] Drishti Yadav, Shilpee Yadav, and Karan Veer. 2020. A comprehensive assessment of Brain Computer Interfaces: Recent trends and challenges. *Journal of Neuroscience Methods* 346 (Dec. 2020), 108918. https://doi.org/10.1016/j.jneumeth.2020.108918

[368] Hiroo Yamamura, Holger Baldauf, and Kai Kunze. 2020. Pleasant Locomotion–Towards Reducing Cybersickness using fNIRS during Walking Events in VR. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 56–58.

[369] Shuo Yan, GangYi Ding, Hongsong Li, Ningxiao Sun, Yufeng Wu, Zheng Guan, Longfei Zhang, and Tianyu Huang. 2016. Enhancing Audience Engagement in Performing Arts Through an Adaptive Virtual Environment with a Brain-Computer Interface. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. Association for Computing Machinery, New York, NY, USA, 306–316. https://doi.org/10.1145/2856767.2856768

[370] Dong Yang, Wei-Xin Ren, Yi-Ding Hu, and Dan Li. 2015. Selection of optimal threshold to construct recurrence plot for structural operational vibration measurements. *Journal of Sound and Vibration* 349 (Aug. 2015), 361–374. https://doi.org/10.1016/j.jsv.2015.03.046

[371] Qi Yang, Shuo Feng, Tianlin Zhao, and Saleh Kalantari. 2023. Co-Design with Myself: A Brain-Computer Interface Design Tool that Predicts Live Emotion to Enhance Metacognitive Monitoring of Designers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3544549.3585701

[372] Yu-Xuan Yang, Zhong-Ke Gao, Xin-Min Wang, Yan-Li Li, Jing-Wei Han, Norbert Marwan, and Jürgen Kurths. 2018. A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28, 8 (Aug. 2018), 085724. https://doi.org/10.1063/1.5023857

[373] Jong Chul Ye, Sungho Tak, Kwang Eun Jang, Jinwook Jung, and Jaeduck Jang. 2009. NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage* 44, 2 (2009), 428–447. https://doi.org/10.1016/j.neuroimage.2008.08.036

[374] Byoung-Kee Yi and Christos Faloutsos. 2000. Fast time sequence indexing for arbitrary Lp norms. (2000).

[375] Catherine Yue. 2023. Privacy and Ethical Concerns of Brain-Computer Interfaces. In *2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom)*. 134–138. https://doi.org/10.1109/MetaCom57706.2023.00036

[376] Beste F. Yuksel, Michael Donnerer, James Tompkin, and Anthony Steed. 2010. A Novel Brain-computer Interface Using a Multi-touch Surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 855–858. https://doi.org/10.1145/1753326.1753452

[377] Beste F. Yuksel, Kurt B. Oleson, Lane Harrison, Evan M. Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. 2016. Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5372–5384. https://doi.org/10.1145/2858036.2858388

[378] Chaw Thet Zan and Hayato Yamana. 2016. An improved symbolic aggregate approximation distance measure based on its statistical features. In *Proceedings of the 18th international conference on information integration and web-based applications and services*. 72–80.

[379] Thorsten O Zander, Jonas Brönstrup, Romy Lorenz, and Laurens R Krol. 2014. Towards BCI-based implicit control in human–computer interaction. In *Advances in Physiological Computing*. Springer, 67–90.

[380] Thorsten O. Zander, Marius David Klippel, and Reinhold Scherer. 2011. Towards Multimodal Error Responses: A Passive BCI for the Detection of Auditory Errors. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. ACM, New York, NY, USA, 53–56. https://doi.org/10.1145/2070481.2070493

[381] Thorsten O. Zander and Christian Kothe. 2011. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems

in general. *Journal of Neural Engineering* 8, 2 (March 2011), 025005. https://doi.org/10.1088/1741-2560/8/2/025005

[382] Thorsten O Zander, Christian Kothe, Sabine Jatzev, and Matti Gaertner. 2010. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-computer interfaces*. Springer, 181–199.

[383] Joseph P. Zbilut and Charles L. Webber. 1992. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A* 171, 3 (Dec. 1992), 199–203. https://doi.org/10.1016/0375-9601(92)90426-M

Appendix A

Detailed Experiment Model with Examples

In the tables below, we present the experiment model for HCI research with brain signals, described in Section 3.6. We describe each category and list the attributes contained within them. For each attribute, we give a definition of the attribute and present an example taken from the surveyed literature. We tried to identify examples which are biased towards a more detailed documentation, but individual examples may still lack certain information. An online version which allows to submit changes is available at `https://brain-signals-hci.github.io/experiment-model/`.

Aspect of Model	Example
Technical Aspects of Recording	
Type of Sensor: For a given brain sensing modality, report the manufacturer and the specification of the sensor chain employed.	*"The EEG was recorded using a NeuroScan system with 32 Ag-AgCl electrodes" (Lee et al., [194])*
Sensor Position: Report where on the scalp electrodes are positioned. For EEG, this is most often done in terms of the 10-20 positioning system or its refinements. For fNIRS, the placement of transmitters and receivers has to be distinguished and the respective distances need to be reported.	*"Electrodes were positioned according to the extended 10-20 system on CPz, POz, Oz, Iz, O1 and O2" (Evain et al., [86])*
Sampling Rate: Report the number of samples recorded per second in Hz.	*"A sampling rate of EEG signals was set as 300 Hz." (Terasawa et al. [323])*
Measurement Quality: For EEG, the threshold for the maximum impedance level (in $k\Omega$) is often reported. For fNIRS, no standardized quality measurement exists, different devices provide different ways of measurement (e.g. photon count).	*"electrode impedance was below 5 $K\Omega$" (Vi et al. [336])*
Reference: Specific to the EEG signal, it is custom to report the electrode to which the recording was referenced.	*"Two electrodes were located at both earlobes as reference and ground." (Terasawa et al. [323])*
Auxiliary signals: The brain sensing modality may not be the only signal, which is captured during the experiment. Often, other sensors, such as eye trackers or heart rate monitors are employed. For these, similar information as for the brain sensing modality may be reported, especially about the specific type of sensor and its placement.	*"Eye positions were measured with an embedded infrared eye-tracking module: aGlass DKI from 7invensun (https://www.7invensun.com)." (Ma et al.[213])*
Synchronization with stimuli and other signals: For analyzing a continous stream of brain signal data, it needs to be synchronized to the events of the experiments (and potentially any other signal sources). This can be done through timestamps, trigger signals, light sensors or other means and the method may be reported to determine the precision of the achieved synchronization.	*"A parallel port connection between recording PC and experimental PC synchronized the EEG recording with the experimental events, such as the sound onset and button press." (Glatz et al., [107])*

Table A.1: Experiment Model: Part 1.

Aspect of Model	Example
Recording Environment: This point reports where and under what conditions the experiment was conducted. Of relevance can be the conditions regarding control of light, sound, electromagnetic fields as well as the positioning of the participant. This attribute is often illustrated through a photo or video of the environment.	*"#Scanners was presented in an intimate 6 person capacity cinema, within a caravan [...]. The space had no windows, low lighting, plush seating, an eight foot projected image, and stereo speakers. Figure 3 [shows a] participant wearing the EEG device, experiencing #Scanners inside the caravan." (Pike et al.[266])*
Task Description	
Participant Restraints: This attribute relates to any instructions or physical restraints, which were in place during the experiment to avoid artifacts or other undesired effects influencing the signal.	*"the participants were instructed to refrain from excessive movement by keeping arms at rest on the table in a position that allowed them to reach the keyboard without excessive movement." (Crk et al., [62])*
Output devices: Describes through which devices (e.g. computer screen, mobile phone, etc.) information and material is communicated to the user. As many brain activity patterns are sensitive to the specific characteristics of the stimulation, details of the presentation may be reported.	*"The [...] game stimulus was run on a powerful high-end gaming PC (CPU: Intel® Core ™ i7-6850K @ 3.60 GHz; RAM: 32 GB; GPU: NVIDIA Geforce GTX 1080) and displayed on a 27-inch BenQ ZOWIE XL2720 144 hz gaming monitor at a 1920x1080 resolution." (Terkildsen and Makransky,[324])*
Input devices: Describes through which devices (besides the brain signal itself) the user communicates commands and other types of input to the system.	*"HMD-mounted Leap Motion (https://www.leapmotion.com) to track participants' hands." (Škola and Liarokapis [345])*
User Input: Describes which kind of commands and input users can enter into the system at which point of the task. May specify which input devices are used and whether there are any requirements or restrictions to the input.	*"They had to respond to auditory notifications whenever one was presented, with a button press using either their left or right index fingers. Six notifications (i.e., 3 complementary pairs of verbal commands and auditory icons) were pre-assigned to a left index-finger press and the remaining six, to a right index-finger press." (Glatz et al., [107])*
Middleware/Communication: For interactive applications or distributed recording setups, this attribute reports how the different parts communicate to exchange data, triggers, commands, etc.	*"We wrote a custom Java bridge program to connect the headset to the Android OS and Unity application on the Game tablet. The Java program polled the headset 60 times a second for EEG power spectrum [...] We connected the Calibrate tablet to the Game tablet using WiFi Direct [. . .]." (Antle et al. [10])*

Table A.2: Experiment Model: Part 2.

Aspect of Model	Example
Framework/Technical platform: What software or development toolkit (in what version) was used as the foundation to implement the task (e.g. PsychoPy, Unity, etc.)	*"The scene was developed using Unity version 2017.3.0f3, for the representation of hands, the realistically looking hand models "Pepper Hands" from Leap Motion suite were used (visible in Figure 3)." (Škola and Liarokapis [305])*
Task Functionality: Reports what functionality the involved software provides to the user (in the case of a working interactive application) and how it responds to different user input. For experiments which are based on or inspired by established paradigms (e.g. from cognitive psychology), this source may be reported (e.g. in reference to a source such as the Cognitive Atlas [268].	*"the main task for the study is a multi-robot version of the task introduced in [27]. Participants remotely supervised two robots (the blue robot and the red robot) that were exploring different areas of a virtual environment. Participants were told that the two robots had collected information that needed to be transmitted back to the control center. [continues. . .]" (Solovey et al.,[307])*
Architecture: For experiments which involve non-trivial custom software artifacts, this attribute reports the underlying software architecture, informing about structure of and information flow between modules.	
Stimulus Material: For tasks which involve the repeated presentation of uniform stimuli (e.g. pictures to rate, text prompts to enter, etc.), this attribute reports the form of these stimuli (e.g. picture size, length, language, etc.) and their source.	*"To prepare experimental materials, a dataset of notifications from the websites Notification Sounds and Appraw were collected. Seven musically trained raters were recruited to determine the melody complexity of the 40 notifications. [. . .] (The stimuli can be downloaded at https://goo.gl/SnZrzG)." (Cherng et al.,[53])*
Visualization provided?: This attribute reports visually (through screen shots or video) the task as shown to the user. If the task has multiple distinct parts, all of them may be visualized. If the task is not in English, the visualization may be accompanied by a translation.	
Timing: Reports on if and how the task is (partially) paced by an internal clock, for example for controlling the duration for stimulus presentation or the time time for responding to a prompt.	*"Each trial began with a black screen for 3s, followed by a fixation dot in the center of the screen for 200ms. After that, the screen remains clear for 200ms before one of four stimuli was displayed for 300ms." (Vi et al. [336])*
Code for task provided?: Reports on whether the task is provided in source code or an executable file and under what licence. If custom hardware is involved, this could also include a blueprint or a circuit diagram.	

Table A.3: Experiment Model: Part 3.

Aspect of Model	Example
Participants	
Recruitment strategy: How where study participants recruited, e.g. through social media, in class, etc.?	*"A snowball procedure was used to gather the sample of study participants. The study was advertised via university courses, email and social media." (Johnson et al.,[159])*
Incentives: What compensation (if any) was offered to study participants, e.g. money, class credit, etc.? What were the criteria for being eligible for the compensation?	*"Participants received monetary compensation for their participation (10 Euro)." (Putze et al.,[278])*
Age: How old are the participants (mean and standard deviation)?	*"mean age* 24.53 *(SD:* 3.00*)" (Frey et al.,[99])*
Gender: With what gender do participants identify (relative frequencies)?	*"2 females and 9 males" (Ma et al.,[213])*
Occupation: What is the profession or - in case of students - the field of study of the participants?	*"Data were collected from 34 computer science undergraduates at the first two authors' institution" (Crk et al., [62])*
Inclusion or exclusion criteria: Where there rules on which participants were eligible to take part in the experiment and what were these criteria (e.g. handedness, disabilities, caffeine consumption, etc.)?	*"Each of the individuals was enrolled in at least one computer science course" (Crk et al., [62])*
Approval of ethics committee: Was the study approved by an ethics committee? If so, by which one?	*"the experimental protocol was approved by the University Research Ethics Committee prior to data collection." (Burns and Fairclough, [39])*
Experiment Flow	
Experiment Structure: Order of different segments of the experiment, such as instruction, sensor placement, training, debriefing, etc. For "in-the-wild" experiments, which do not follow a fixed, predefined pattern, this attribute may report the boundary conditions and which parts of the experiment could be.	*"•5-10min of explanation (slideshow with a detailed description of the interface), •5min to set up the hardware, •For each BCI paradigm: 10min of calibration, familiarization •10min to play the game •5min of rest •Questionnaires" (Kosmyna et al.,[179])*
Instructions: Instructions given to the participants regarding purpose of experiment, experiment setup, task operations, restrictions, safety considerations, etc. If possible, the written instruction documents may be provided.	*"Participants were instructed to focus visual attention on a target symbol, whilst silently counting the number of times the target character flashed." (Obeidat et al.,[254])*

Table A.4: Experiment Model: Part 4.

Aspect of Model	Example
Training procedure: This attribute reports about how the participants familiarized themselves with the task. it may report the duration of training, specific training conditions compared to the main experiment, and specific training instructions.	*"All participants first undertook a training task where they played 15 easy and 15 hard pieces on the piano. Each piece was 30 seconds long with a 30-second rest between each piece." (Yuksel et al.,[377])*
Trial ordering: For any experiment that is organized in trials, this attribute reports how the ordering of the blocks was derived. It may report whether and how ordering differed between participants.	*"The order of conditions was counter-balanced across subjects and participants wore the fNIRS device during both trials so they did not notice a difference between the two conditions." (Afergan et al.,[4])*
Repetitions, blocks & breaks: This attribute reports the larger structure of the experiments, such as blocks of trials and their duration, ordering as well as pauses between blocks. If applicable also reports stopping criteria, if these are individual.	*"Each session comprised four testing blocks: two using club ambient noise and two using city street ambient noise as standard stimuli. Between any two consecutive blocks, subjects had three minutes to rest." (Lee et al.,[194])*
Pre-study screenings: Reports any procedures prior to the experiment that determine the eligibility of a participant for the study, their assignments to an experiment group or other aspects of the experiment process.	*"None of the subjects had any history of brain disease, drug use, or hearing problems. None had any musical expertise." (Lee et al.,[194])*
Questionnaires: Reports which questionnaires were administered before, during and after the experiment. May give a reference to a published questionnaire or list the items of a custom one. May also report when and how often questionnaires were administered and through what means (paper or computer-based).	*"At the end of the evaluation, end-users were asked to complete the NASA-TLX, the eQUEST 2.0, and a customized usability questionnaire. [. . .] After each session was complete, the therapist station would automatically open up on the laptop and ask the user to answer, How satisfied were you with the BCI session? (with 10 being very satisfied and 0 is not satisfied) [continues . . .]" (Miralles et al., [234])*
Data Processing	
Derivation of labels: Outside neurofeedback applications the recorded brain signal data is distributed between multiple groups or assigned a continuous value. This attribute may report how the label is derived from the collected data (e.g. defined by the experiment structure, by questionnaire responses, or external ratings).	*"We considered the mean of the three NASA-TLX parameters (effort, mental demand and frustration) to evaluate the overall mental workload. The average score was thresholded at the mean value of 2 (since the used scale was 0–4) to quantize or characterize a parameter block as inducing low/high workload." (Bilalpur et al.,[29])*

Table A.5: Experiment Model: Part 5.

Aspect of Model	Example
Data transformation: This attribute refers to all processing steps which transform raw data while keeping it in the original time-domain representation. Examples of such transformation steps are: re-referencing, baseline normalization, downsampling, etc.	*"the common average was subtracted from all EEG channels." (Lampe et al.,[192])*
Filtering: This attribute reports any filtering of the data. This may include the type of filter applied as well as necessary parameters, such as the filter order.	*"EEG data was first low-pass filtered with a cutoff frequency of* $50hz$ *and high-pass filtered with a cutoff frequency of* $0.16hz$*, both using a third-order butterworth filter" (Rodrigue et al.,[288])*
Windowing: This attribute reports how segments of data are aligned (e.g. locked to an event in the experiment), how long they are and with which window function they are extracted.	*"The data was then segmented into 1.5-second epochs, overlapping each previous epoch by* 50%*" (Rodrigue et al.,[288])*
Artifact cleaning: Reports through which algorithms (beyond filtering) artifacts were removed and which artifacts are targeted.	*"Independent Component Analysis (ICA) is applied [...] The components are first filtered using a band-pass filter with cut off frequencies 1 - 6 Hz. Choosing the component with the highest energy [and] applying a high-pass filter with a cut off frequency of 20 Hz." (Jarvis et al., [155])*
Hyperparameter optimization: For machine learning models, this attribute reports how the hyperparameters of the model were chosen (e.g. through grid search) and which hyperparameters where chosen in the final model. This also includes other parameters of the processing pipeline which are optimized (e.g. in preprocessing).	*"A grid search was performed to optimize sigma for all participants, the remaining parameters were left as default." (Rodrigue et al.,[288])*
Outlier handling: Reports any methods for excluding certain samples, windows, or sessions based on the contained data or other external factors.	*"Any rest or trial period with 20 percent or higher error rate is considered noisy and can be excluded from the analysis" (Crk et al., [62])*
Feature extraction: This attribute reports on how a feature vector for classification or regression is calculated from the preprocessed data.	*"[W]e partitioned each data window into smaller segments of 50 ms length. We then used the signal mean of the segment, calculated on the band-pass filtered signal, with cutoff frequencies at 4 and 13 Hz (i.e.* θ*- and* α*-bands)." (Putze et al., [278])*

Table A.6: Experiment Model: Part 6.

Aspect of Model	Example
Feature selection: This attribute reports on procedures to reduce the number of features automatically.	*"we performed a feature selection using the Fisher ratio as selection criterion. The number k of selected features [...] was a tuning parameter in the range between 5 and 50." (Putze et al., [278])*
Learning model: This attribute reports on the specific machine learning model that is employed (if any) to perform classification or regression.	*"the Neural Network Toolbox of MATLAB was used to create an artificial neural network (ANN) with 198 inputs, 20 hidden neurons and 4 outputs. The patternnetfunction, which creates a feed-forward neural network, was used. [...]" (Lampe et al.,[192])*
Evaluation procedure: For machine learning models, this attribute reports how they were evaluated to assess their performance. This involves the exact metric used for assessment as well as the approach to (sometimes repeatedly) determine test and training data sets.	*"To assess the classifiers' performance on the calibration data, we used 4-fold cross-validation (CV). [...] The performance was measured using the area under the receiver-operating characteristic curve (AUROCC)." (Frey et al.,[99])*
Processing code provided? Reports if the code for processing the brain signal data is released with the paper or in a separate repository. If the code cannot be provided, as a substitute it is possible to report the employed frameworks (e.g. EEGLAB).	*"The full classification pipeline is implemented in Python. For EEG processing, we use the MNE toolbox [17]. For machine learning and evaluation algorithms, we use scikit [28] and custom routines build on numpy and scipy." (Putze et al.,[278])*
Brain Signal Integration	
Brain Input effect: This attribute describes how the output of the brain input processing influences the design, the behavior, or the content of the application or experimental paradigm.	*"when the system was confident that the user was in a state of low or high workload, one UAV would be added or removed, respectively. After a UAV was added or removed, there was a 20 second period where no more vehicles were added or removed." (Afergan et al.,[4])*
Type of integration: This attribute describes the algorithmic implementation of the brain signal integration, i.e. whether an explicit conditional statement, an Influence Diagram, a state graph, or a different way of behavior modeling was used.	*"[The self-correction algorithm] inspects the probability distribution [...] and picks the now highest scoring class [...]. [W]e only used the second best class if its re-normalized confidence [...] is above a certain threshold T [...]. Otherwise, the user was asked to repeat the input." (Putze et al.,[273])*

Table A.7: Experiment Model: Part 7.

Appendix B

Team Collaboration Task Demographics

Here we provide demographic information for participants in the escape room design team collaboration task from Chapter 6.

Table B.1: Team collaboration task participant demographics (Part 1). Escape room familiarity levels could range from (1 - Not at all familiar) to (4 - Very familiar), as in Section 6.2.2; the final column lists the location of each participant (Location 1 - WPI or Location 2 - Uni-Bremen).

ID	Age Range	Gender	Ethnicity	Education	Familiarity	Group	Location
P1	18-24	Male	White	High school	4	1	Location 1
P2	18-24	Male	Asian	College	3	1	Location 1
P3	18-24	Male	Asian	Graduate degree	2	1	Location 2
P4	25-34	Male	Asian	Professional degree	2	1	Location 2
P5	18-24	Male	White	High school	2	2	Location 1
P6	18-24	Male	Asian	Professional degree	3	2	Location 1
P7	25-34	Male	Asian	Professional degree	2	2	Location 2
P8	25-34	Female	White	Graduate degree	2	2	Location 2
P9	18-24	Male	Asian	Graduate degree	2	3	Location 1
P10	18-24	Male	Asian	High school	2	3	Location 1
P11	35-44	Female	White	College	3	3	Location 2
P12	18-24	Male	Asian	High school	1	3	Location 2
P13	18-24	Male	White	College	3	4	Location 1
P14	18-24	Female	Asian	Professional degree	1	4	Location 1
P15	18-24	Female	Other (Amazigh)	High school	3	4	Location 2
P16	18-24	Female	White	College	2	4	Location 2
P17	18-24	Male	Asian	Graduate degree	4	5	Location 1
P18	18-24	Other	Asian	High school	4	5	Location 1
P19	25-34	Female	Asian	Professional degree	3	5	Location 2
P20	18-24	Male	Asian	College	3	6	Location 1
P21	18-24	Male	White	High school	3	6	Location 1
P22	18-24	Male	Asian	Graduate degree	2	6	Location 2
P23	25-34	Female	Asian	Graduate degree	1	6	Location 2
P24	25-34	Male	Asian	College	2	7	Location 1
P25	18-24	Male	White	High school	3	7	Location 1
P26	18-24	Male	Asian	Graduate degree	2	7	Location 2
P27	18-24	Male	Other (Indian)	Graduate degree	2	7	Location 2
P28	18-24	Male	White	High school	4	8	Location 1
P29	25-34	Male	Asian	Professional degree	2	8	Location 1
P30	18-24	Female	White	High school	3	8	Location 2
P31	18-24	Female	Asian	College	3	8	Location 2
P32	25-34	Male	White	Graduate degree	4	9	Location 1
P33	25-34	Female	White	Graduate degree	3	9	Location 1
P34	18-24	Female	Asian	Graduate degree	1	9	Location 2
P35	25-34	Male	Asian	College	2	10	Location 1
P36	25-34	Male	Asian	Professional degree	2	10	Location 1
P37	25-34	Female	Asian	Professional degree	1	10	Location 2

Table B.2: Team collaboration task participant demographics (Part 2). Escape room familiarity levels could range from (1 - Not at all familiar) to (4 - Very familiar), as in Section 6.2.2; the final column lists the location of each participant (Location 1 - WPI or Location 2 - Uni-Bremen). Missing information from two participants who did not complete the demographic questionnaire is omitted.

ID	Age Range	Gender	Ethnicity	Education	Familiarity	Group	Location
P38	25-34	Female	Other (Middle Eastern)	Graduate degree	2	11	Location 1
P39	18-24	Male	Asian	Graduate degree	3	11	Location 1
P40	-	-	-	-	1	11	Location 2
P41	18-24	Male	Other (Mixed, White and Asian)	High school	3	12	Location 1
P42	18-24	Non-binary	White	High school	3	12	Location 1
P43	-	-	-	-	2	12	Location 2
P44	25-34	Female	White	College	4	12	Location 2