Capturing Team Cognition: A Multimodal Dataset for Adaptive Collaborative Interfaces

Christopher Micek
Worcester Polytechnic Institute
Worcester, Massachusetts, USA
cjmicek@wpi.edu

Lasse Warnke
University of Bremen
Bremen, Germany
lwarnke@uni-bremen.de

Lourenço Abrunhosa Rodrigues
University of Bremen
Bremen, Germany
lourenco@uni-bremen.de

Felix Putze
University of Bremen
Bremen, Germany
felix.putze@uni-bremen.de

Erin Solovey
Worcester Polytechnic Institute
Worcester, Massachusetts, USA
Harvard-Radcliffe Institute
Cambridge, Massachusetts, USA
esolovey@wpi.edu

Abstract

We introduce a multimodal dataset and experimental setup designed to support the development of adaptive collaborative systems. Data were collected from distributed teams working simultaneously across two continents, demonstrating the feasibility of sensing team cognition in geographically dispersed settings. The dataset includes synchronized EEG, audio transcripts, screen recordings, and behavioral annotations, enabling fine-grained analysis of collaboration in naturalistic settings. Our setup integrates neural and behavioral sensing to model team processes, using metrics such as task engagement, neural synchrony, and interaction patterns. These analyses reveal relationships between cognitive states and team dynamics, suggesting new directions for brain-computer interfaces that respond to team-level signals. By providing a shareable dataset, robust sensing infrastructure, and techniques for modeling distributed collaboration, this work enables future interactive systems that sense and support distributed teamwork in real time.

CCS Concepts

• **Human-centered computing → Collaborative and social computing design and evaluation methods**.

Keywords

brain-computer interface, teamwork, collaboration, EEG, hyperscanning

ACM Reference Format:
Christopher Micek, Lasse Warnke, Lourenço Abrunhosa Rodrigues, Felix Putze, and Erin Solovey. 2026. Capturing Team Cognition: A Multimodal Dataset for Adaptive Collaborative Interfaces. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3772318.3791607

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791607

1 Introduction

Teams often struggle to maintain shared understanding, coordinate actions, and stay engaged [20, 64, 73, 84], and these difficulties are amplified when members are geographically dispersed [16, 54]. While digital collaboration tools have improved communication and task management [85, 112], they provide little visibility into the underlying cognitive and affective states that shape team performance. This lack of awareness can lead to breakdowns in coordination, reduced creativity, and lower overall effectiveness. Consequently, effectively monitoring team dynamics to assess function, predict breakdowns, and trigger timely interventions has been a longstanding concern in human factors and organizational psychology [61, 98].

Assessing these dynamics has traditionally relied on self-report measures administered after or during collaboration, which can be biased by team members' subjective recollections and disrupt team workflow [71, 75, 119]. Furthermore, such assessment only captures momentary snapshots of the relevant processes, rather than providing precise, moment-by-moment measurements. Although there has been much work examining team interactions through manual annotation of audio or video recordings, this approach can be onerous for researchers, and may not scale to the needs of large organizations [14, 72, 119]. In contrast, neurophysiological signals can be collected continuously and unobtrusively, [113], while additionally offering a more mechanistic understanding of how their evolution differs for successful versus unsuccessful collaboration [45, 113]. Physiological measures can therefore complement or, in some situations, more accurately represent key collaborative processes than post-hoc subjective assessments [55].

Recent advances in physiological computing and brain-computer interfaces (BCIs) offer a new opportunity to leverage these benefits by sensing cognitive and behavioral states in real time to inform adaptive support [49, 104, 123]. However, most BCI research has focused on individual users in controlled settings [2], leaving open questions about how to capture and interpret neural signals in dynamic, multi-person, distributed contexts. Existing studies have primarily examined co-located pairs, with fewer incorporating combined neural and behavioral measures from groups beyond dyads [93]. Applying BCIs to group creativity in real-time collaborative settings remains underexplored. In addition, recording extensive, high-quality EEG data from multiple people interacting simultaneously in a distributed setting is both time-consuming and technically complex [60, 66]. As a result, there are few publicly available datasets that enable modeling of team cognition in realistic, geographically distributed collaboration.

This paper addresses these gaps by introducing a multimodal dataset and experimental framework tailored to studying team cognition in remote creative collaboration. We present a novel study design that captures synchronized EEG, audio transcripts, screen interactions and behavioral annotations from remote teams engaged in a complex creative task across two continents.

Building on this dataset, we explore methods for linking neural and behavioral signals to team processes and performance. Specifically, we calculate two individual-level EEG measures—the Task Engagement Index (TEI) and Task Load Index (TLI)—as well as measures derived from four group-level analysis methods: multidimensional recurrence quantification analysis (MdRQA), mutual information (MI), wavelet transform coherence (WTC), and the driver-empath model of team dynamics (DE). We also analyze transcripts of the experiment using Marks et al.'s taxonomy of team processes [71] to extract behavioral indicators. These measures are evaluated in relation to team performance and team processes using techniques such as windowed correlation and regression analysis.

By providing a shareable dataset, robust sensing experiment infrastructure and analysis techniques, this work lays the foundation for adaptive interfaces that monitor and enhance distributed teamwork. The contributions of this paper are as follows:

(1) **Advancing BCIs from individual to team-level applications for distributed, real-time collaboration:** We demonstrate the feasibility of capturing synchronized EEG and behavioral data to monitor cognitive states in distributed teams engaged in creative tasks across two continents. This work shifts the focus from individual monitoring toward team-level cognitive and behavioral dynamics, an area that remains relatively underexplored in HCI.
(2) **Introducing a multimodal dataset for studying team cognition and interaction in naturalistic settings:** Our dataset is derived from synchronized EEG, video, audio, screen recordings, and behavioral data from distributed teams engaged in a complex creative task. This resource supports the development and testing of real-time supportive user interfaces for distributed teams. It enables researchers to assess the ecological validity of neural synchrony and other cognitive metrics as indicators of successful collaboration demonstrated previously in controlled settings [22, 33, 42, 80], and provides a foundation for training and testing real-time monitoring models to drive adaptive interfaces.
(3) **Establishing methods for linking neural signals to team processes and performance:** We integrate individual-level EEG metrics (Task Engagement Index, Task Load Index) with group-level synchrony measures (Mutual Information, Wavelet Transform Coherence, Recurrence Rate, and Driver-Empath Synchrony) as well as behavioral indicators such as speech-based team process coding, offering new techniques for modeling coordination and workload in collaborative systems.
(4) **Providing a foundation for adaptive interfaces that respond to team-level cognitive and behavioral states:** Our analyses reveal modest but interpretable relationships between neural activity and team behavior. Notably, mutual information and recurrence rate show associations with task-focused and interpersonal processes, respectively, suggesting that group-level neural synchrony may reflect key aspects of team functioning. These metrics offer a potential pathway for lightweight, scalable BCI systems that detect breakdowns in collaboration and deliver timely support.

2 Background

2.1 Foundations of Team Cognition and Processes

Teams are complex dynamical systems [74], 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.) [8, 27]. 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.) [12, 71]. Given that the affective states, cognitive states, and behavioral processes of team members are indicative (and sometimes predictive [58, 95]) 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 and minimizing process loss. Traditionally, assessments have relied on administering periodic surveys and behavioral assessments [71, 75, 119], which can disrupt team workflow and only provide sporadic indicators of how well they are collaborating. 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 [84, 108]. This is an area where physiological sensing offers unique promise.

2.2 Neurophysiological Sensing for Collaboration and Creative Problem Solving

Recent advances suggest that collaborative tools 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 [67, 125]. For example, EEG-based indices have successfully been used to detect lapses in attention and cognitive fatigue, supporting their relevance in collaborative scenarios [67]. Prior work has demonstrated that affective states, such as valence and arousal [23, 28, 51], and cognitive states such as mind wandering [25], multitasking [1, 103], learning stages [47], and workload [3, 88, 105] can be measured and integrated into online systems via continuous non-invasive physiological sensing. These approaches offer continuous, unobtrusive monitoring and provide insights into cognitive states, complementing traditional behavioral measures.

This ability to sense and respond intelligently to mental states and processes has led researchers to explore the potential for BCI-driven 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) [31]. A wide variety of such tools have been developed both for individuals [115] and groups [32], including tools which help define problem scope of prior to brainstorming [7], allow users to map the semantic connections between ideas [101], iteratively generate new graphics based on initial user input [52], and more [115]. BCI-based support tools are additionally able to respond to the cognitive and affective states of users. Botrel, Holz, & Kübler [13] and Todd et al. [109] 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 [123], neurocognitive feedback to enhance creative problem solving [49], and a brainstorming assistant that can alter the semantic distance between suggested ideas in response to a user's level of cognitive effort [18].

2.3 Interpersonal Neural Synchrony and Hyperscanning

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 [22, 29]. Various measures including wavelet transform coherence [22], the recurrence rate from multidimensional recurrence quantification analysis [29], and mutual information [106] have been demonstrated to assess joint mental processes during collaboration. Preliminary work employing hyperscanning—in which brain data from multiple users is measured simultaneously—using electroencephalography (EEG) [26, 78] and fNIRS [22, 47] demonstrated differences in neurophysiological signals between participants working individually versus as a team, and between expert and novice teams. For example, by employing fNIRS hyperscanning 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) [29].

2.4 Limitations of multi-user BCI systems

While brain sensing offers solutions to these assessment limitations, the transition to real-world, multi-person support faces two major challenges. First, the majority of BCI systems supporting creative output or ideation have focused on assisting individual users [13, 18, 49, 109, 123]; while existing support tools are able to leverage brain activity to provide novel forms of assistance, they lack the ability to simultaneously record and process the brain data of multiple users. Second, research on inter-brain dynamics itself remains constrained. Although work on multi-user systems dubbed *collaborative BCIs* (cBCIs) exists [10, 110, 116], these systems are often tailored for active control or decision making scenarios, where neural signals are aggregated to improve precision on a limited set of well-defined tasks (e.g., target selection [10, 116], or visual search [110]). While cBCIs exhibit greater classification performance than single-user systems, they are only applicable to the limited set of circumstances where (i) multiple users all perform the same task with little if any interaction, and (ii) the desired output is a classification decision. This contrasts sharply with the need for passive monitoring tools that track complex, open-ended team processes. Consequently, the literature catering to complex, multi-person contexts remains sparse; hyperscanning studies have been predominantly limited to dyads [93], with only a small number examining larger groups.

2.5 Methodological Considerations for Hyperscanning Studies

In addition, hyperscanning studies face challenges such as synchrony driven by shared stimuli or shared knowledge [17, 124] and EEG-specific artifacts like volume conduction [43]. While these factors complicate interpretation of neural coupling as evidence of social cognition, they do not necessarily diminish the value of synchrony as a proxy for team states in applied contexts. Reviews recommend mitigation strategies including spatial filtering, and comparisons to within-brain baselines, and the use of surrogate datasets (derived from permuted subject pairs or phase randomization) to reduce spurious correlations [17, 21, 124]. Other approaches include using asymmetric or independent stimuli across participants, incorporating individual-task baseline conditions, and analyzing lagged rather than zero-lag coherence to reduce the risk of attributing coincident entrainment to social coupling [65, 124].

While these methodological challenges underscore the complexity of interpreting neural coupling, they also highlight the need for standardized, openly available datasets that enable replication, benchmarking, and exploration of mitigation strategies. To address this gap, we next review existing resources and identify opportunities for advancing research on team cognition in realistic collaborative settings.

2.6 Open Datasets and Research Gaps

To situate our work within this landscape, we conducted a broad search for publicly available hyperscanning datasets collected from participants engaging in realistic collaboration tasks. In order to be considered, studies needed to (i) provide an open-access EEG dataset, and utilize an experiment design in which (ii) participants worked together to find creative solutions to a problem or achieve a common goal (iii) in teams of three or more.

We searched OpenNeuro, NeurIPS, and OSF collections using the query "EEG AND (team OR collaboration)", and Google Scholar Labs[1] using the prompt "Find studies providing openly available EEG hyperscanning datasets where participants worked together on teams of three or more for a realistic collaboration task (e.g., design, creative problem solving, search-and-rescue, etc.)." Our initial search yielded 815 records, which we then narrowed to four studies after removing duplicates and applying our inclusion criteria

[1]https://blog.google/outreach-initiatives/education/google-scholar-labs/

(Table 1). Datasets differ in terms of multimodality— with some incorporating information from speech and eye movement in addition to EEG [89, 91]—and the level of creative freedom afforded to team members. While some experiment tasks were largely unconstrained (jazz improvisation [92], brainstorming [94]), most had specific goal states teams needed to reach (search-and-rescue [89], spaceship navigation [91], Sudoku [94]). All but one dataset [91] recorded brain activity from participants co-located in the same room, and none included data from geographically dispersed teams. To date, no openly available resource provides the multimodal data necessary to model team cognition during geographically distributed, complex creative collaboration. However, with the increasing rate of research on brain-computer interfaces in the real-world [87, 104] 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.

These gaps motivate our contribution: a multimodal dataset and experimental framework capturing synchronized EEG and behavioral data from distributed teams engaged in creative collaboration across continents. In the next section, we describe the study design and data collection methods that enable this resource, followed by details of the dataset and its potential for advancing adaptive collaborative systems.

3 Towards Brain-Computer Interfaces to Support Distributed, Real-time Collaboration: Study Design

This exploratory study was a multi-institute collaboration between Worcester Polytechnic Institute (USA) and University of Bremen (Germany), which leveraged behavioral and physiological data collected from team members in two different continents while they completed a complex creative task in order to gain a more comprehensive understanding of collaboration dynamics in realistic scenarios. The task was designed to model requirements of many modern team tasks, in which the group has to find creative solutions to new problems. Often, these tasks do not have a prescribed solution strategy and require multiple iterations of ideation, implementation, evaluation and are thus highly interactive. 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. We report the experiment description in accordance with the experiment model provided by Putze et al. [87].

3.1 Experiment Procedure

Prior to the experiment session, participants were provided an overview of the study goals and procedures and the informed consent form, which included an explanation that anonymized EEG and task data would be shared publicly after the study and could not be withdrawn once published. During the experiment session, participants were again informed about the study goals, and given the opportunity to ask the researchers any questions. After the briefing, participants signed a consent form and the experiment session proceeded. Participants were reminded that participation was voluntary and that they could withdraw from the study at any time.

Sensors were applied to the scalp and data recording was set up (Section 3.3), participants were provided a brief overview of the Miro digital whiteboard interface, and then given directions for the task (Section 3.2). Participants worked in teams of three or four (up to two from each respective institution) throughout a one-hour session. After the task was completed, we removed the sensors, stopped data acquisition, debriefed the participants, and provided monetary compensation for their time.

3.2 Study Task

Our goal was to identify a task that would capture the essence of distributed creative teamwork while meeting several practical requirements. In particular, the task needed to involve complex, open-ended problem-solving and to require integration of diverse ideas as well as coordination across team members to build a shared mental model. Such a task would allow for observable fluctuations in team and individual dynamics, such as distraction, disengagement, or internal reflection, that occur in real-world collaborative work. Prior work has shown that engaging in such creative collaboration tasks can elicit group-level physiological phenomena such as interbrain synchrony [22, 69, 121], which could serve as real-time indicators of the quality of team interactions. At the same time, the task had to be completed within a reasonable time frame, provide a clear measure of success, and remain engaging enough to sustain participant interest throughout the session. Finally, to reflect emerging workplace practices, we wanted a task that would be sufficiently challenging so that AI assistance would be beneficial. This mirrors real-world collaboration environments where AI tools are increasingly embedded [11, 19, 97], helping researchers examine their impact on team processes and neural measures. These considerations guided our choice of a design scenario that is both engaging and multi-faceted.

The task we chose was 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. This activity combines narrative development, puzzle design and user experience considerations, requiring teams to exchange information and align on a coherent concept. We allowed participants to use ChatGPT, a generative AI chat assistant, as a resource during the task.

Participants had one hour to work together, and were required to document their design via a shared Miro digital whiteboard. This duration was sufficient to capture states observed in real-world teams such as distraction, disengagement, or internal reflection, while remaining practical for study participants. A briefing document introduced virtual escape rooms, key design elements, and ChatGPT usage. Using Krekhov et al.'s [63] taxonomy for escape room games as a foundation, we described several considerations shown in Table 2. Finally, we provided the rubric and scoring scheme that would be used to score their escape room design, likewise based on the taxonomy from [63] with a total possible score out of 100 (Table 12 in Appendix A). Participants were advised that they did not need to build or implement their designs; all that was required was that they work together to create a design and document their

Dataset	Number of Participants	Task	Modalities	Available Artifacts
[89]	40 teams of three	Baseline tasks (resting vs. finger-tapping; individual vs. group affective task; competitive vs. cooperative ping-pong); Group simulated search-and-rescue task	EEG, fNIRS, EKG, GSR, gaze, vocalic features	Raw & preprocessed physiological data, behavioral data from baseline tasks, log data from baseline and group tasks, screenshots
[91]	18 teams of three	VR spaceship navigation task	EEG, pupil size, speech	Participant performance, preprocessed physiological and behavioral data, location/action logs
[92]	Three jazz musicians	Jazz improvisation	EEG	Raw EEG data, audio, video
[94]	44 teams of four	Various problem-solving tasks from the McGrath circumplex [76] (brainstorming, typing, Sudoku, etc.), completed individually or as a team	EEG	Raw behavioral data, preprocessed behavioral and EEG data (overall per-task synchrony), analysis code

Table 1: Open-access EEG hyperscanning datasets exploring collaboration in multi-person teams.

Figure 1: Recording setup at one of the sites, showing participants wearing the EEG cap in two different rooms.

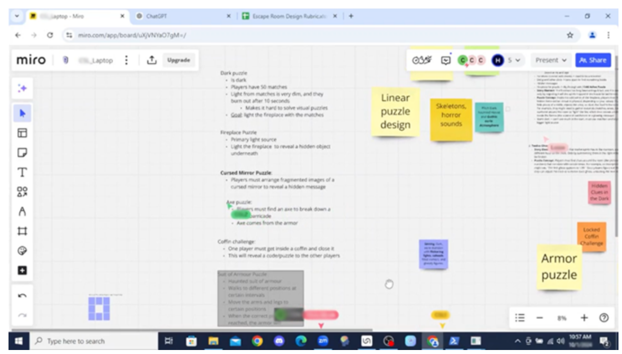

Figure 2: Participant's view of a Miro whiteboard during a recording session.

design process using Miro (Figure 2). 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. 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.

3.3 Distributed, Real-time Data Acquisition

Participation took place remotely, using Zoom to facilitate video and audio recording. Team members were each seated in front of a laptop or desktop computer at their respective locations, all in separate rooms to ensure audio isolation. The isolated lab rooms, exclusively were used for the experiment during that time. Some, but not all, were fitted with electromagnetic and sound absorbing shielding. The computers were either PCs or a laptop running the Windows operating system, and were equipped with a mouse, a webcam, and speakers (Figure 1). 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 4o, and a copy of the rubric that would be used to score designs. Researchers remained on the Zoom call with participants to answer any questions, but turned off their audio and video feeds to avoid distraction. Each participant's camera feed and screen, along with microphone and system audio, were recorded simultaneously using Open Broadcaster Software (OBS) Studio [5]. To ensure the audio and video could be aligned with recorded brain data, we also integrated an OBS plugin[2] which allowed relevant timing information to be published to a data stream in Lab Streaming Layer (LSL) [60], an open source middleware framework for

[2]https://gitlab.unige.ch/sims/lsl-modules/obs-plugin

Table 2: List of design considerations drawn from Krekhov et al. [63] to introduce escape room games and spur design ideas.

Puzzle Types & Design Considerations	Description
Mental Challenges	**Include mental challenges** such as puzzles involving observation, pattern recognition, calculation, and knowledge.
Physical Challenges	**Incorporate physical challenges** that require object movement, alignment, or agility.
Emotional Challenges	**Add emotional challenges** that evoke strong emotions (e.g., unease, fear, surprise), require difficult decisions, or deal with negative consequences.
Target Group & Team Composition	**Specify the number of players** the challenge is made for.
Theme and Narrative	**Define the narrative** by explaining why players are trapped and what **theme or storyline** drives the experience.
Hint System	**Determine hint availability** for players who become stuck.
Failure Handling	**Set a time limit** for completing the escape.

streaming, receiving, and synchronizing multimodal time series. 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.

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 [39]). 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 [82]) 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 band power 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[3] was used to publish LSL streams for each Unicorn device containing their respective channel data.

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[4]. All computers were connected to a common virtual private network (VPN) using the ZeroTier VPN client[5] to ensure that all streams were visible to the computers at both sites.

4 Multimodal Dataset for Studying Team Cognition and Interaction in Naturalistic Setting

Here we summarize our dataset capturing brain and behavioral measures of teams during collaboration, which could be used to further the development of adaptive BCIs to support teamwork. Our dataset and documentation are hosted on the Open Science Framework (OSF), accessible at: https://osf.io/p2u85/overview. A summary of the contained artifacts can be found in Table 3.

[3]https://pypi.org/project/pylsl/
[4]https://github.com/labstreaminglayer/App-LabRecorder
[5]https://www.zerotier.com/

4.1 Participants

Participants were 44 members of the two university student bodies with no known neurological conditions (median age range 18–24, 26 male; full demographics in Table 13). 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). We administered a brief pre-experiment survey via Qualtrics to collect demographics data, measure participants' familiarity with escape rooms, and assess differences in their emotion regulation strategies using the Process Model of Emotion Regulation Questionnaire (PMERQ) [83] and these are summarized in Table 14 in Appendix B. Participants were compensated \$15/hr or 15€/hr, depending on their respective locations, for their time. Recruitment and experimental procedures were approved by both universities' institutional review board or ethics committee. Participants completed informed consent forms upon arriving at each lab.

4.2 Task Performance Data

Task performance was scored using the rubric in Table 12. Two researchers first individually scored each group's designs as documented on the Miro board (referring to the session videos if necessary for clarification), providing numeric scores and notes explaining the reasoning for the score choice for each of the 12 design criteria. If the researchers' scores for any criteria differed by more than one ranking, 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. The final score was calculated by averaging the two researchers' scores for each criterion and summing them.

Overall, team performance varied considerably, with scores ranging from 49.5 to 93.5 out of a possible maximum of 100 ($M = 74.8$, $SD = 14.1$) (Table 4). While there were some groups at the extreme ranges of performance, most groups performed modestly well, reaching scores of ~75 or above.

Table 3: Dataset schema with the included artifacts. A more detailed breakdown can be found in the wiki of the OSF repository.

Artifact	Content	Schema
EEG raw data	8 channels of EEG synchronized for 3–4 participants	$N \times M$ matrix per group, where N is the number of data points in the longest EEG recording, and M = 8 channels × # team members
EEG preprocessed data	Filtered EEG data truncated to task duration	Task duration × M matrix per group
Derived Individual-Level EEG Metrics	Power spectral density, task engagement (TEI) and task load (TLI) indices	1 s non-overlapping segments per metric per participant
	Driver and empath scores	30 s half-overlapping segments per metric per participant
Derived Group-Level EEG Metrics	19 measures from four hyperscanning analysis methods	30 s half-overlapping segments per measure per group
Transcripts	Automatically generated speech transcripts	One line per utterance
Task performance	Scores of resulting escape rooms	12 rubric criteria scores per group (Table 12 & Section 4.2)
Person-specific traits	Demographic information & PMERQ scores	16 data points per participant
Derived Team Process Metrics	Annotation based on vocabulary from [72]	30 s half-overlapping segments per speaker per group (Figure 3 & Section 4.3.2)

Table 4: Escape room design scores (0 to 100) and prior familiarity with escape rooms for all groups. The minimum score (49.5) maps to red, and the maximum (93.5) maps to green, with intermediate values interpolated. Familiarity levels are provided for each participant (P1 or P2) at each study location: 1 - Not at all familiar (red); 2 - Somewhat familiar; 3 - Moderately familiar; 4 - Very familiar (green). Blank cells indicate groups with fewer participants.

Group	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

4.3 Communication and Interaction Data

4.3.1 Raw Communication and Interaction Data. We recorded audio and video data, but due to the challenges of anonymizing audio and video, this data is not part of the public dataset. Instead, we created transcripts, which were then verified for anonymity. In addition, the videos contain the screen content throughout the whole session, including Zoom windows, the shared Miro board, ChatGPT and any other online tools that were used by the team.

We used WhisperX [6] 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).

4.3.2 Derived Team Process Metrics. From participant transcripts, we extracted measurements for 10 dimensions based on the taxonomy of team processes by Marks et al. [71] (Figure 3), a widely utilized framework [14, 29, 59, 72, 96, 120], plus an additional measure of general teamwork. In order to quantify each of these team process dimensions without the need for time-consuming manual annotation, we adopted an automated approach previously used by [72], who employed a validated word and phrase dictionary aligning with Marks et al.'s taxonomy. Each dictionary word could be categorized as one of the 10 lower-order dimensions, one of three higher-order dimensions (if it linked to several lower-order dimensions), or general teamwork (if it connected to multiple higher-order dimensions).

We counted the frequencies of words belonging to each category in successive windows with overlap; composite measures for the

Transition processes
– Mission analysis, formulation, and planning
– Goal specification
– Strategy formulation

Action processes
– Monitoring progress toward goals
– Systems monitoring
– Team monitoring and backup behavior
– Coordination

Interpersonal processes
– Conflict management
– Motivation and confidence building
– Affect management

Figure 3: Team process dimensions based on Marks et al.'s taxonomy [71]. See the original publication for detailed definitions.

transition, action, and interpersonal dimensions were calculated by summing their counts with those of their respective lower-order dimensions. We looked at 30 second windows with 15 seconds of overlap, chosen to match the approximate timescale of meaningful interaction episodes between group members.

This resulted in individual time series for all 44 group members depicting the levels of the higher- and lower-order team processes over the course of the session, as well as for each of the 12 groups when the time series of its members were averaged. Each point in the time series represented one of the 30-second windows. Table 5 summarizes the higher-order team processes for each group.

4.4 Neural Data

EEG data was collected following the procedure outlined in Section 3.3. This section describes the preprocessing of the raw data and derivation of several individual-level EEG measures, starting with power spectral density, and then calculating measures of engagement and workload. To measure the collective brain dynamics of group members, we also detail 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: multidimensional recurrence quantification analysis (MdRQA) [29], mutual information [106]; wavelet transform coherence (WTC) [22, 43] and the driver-empath model of team dynamics [42]. The provided dataset includes all individual and group neural measures as described below.

4.4.1 Raw Data. The data for each participant consist of electrical activity recorded at 250 Hz from positions Fz, C3, Cz, C4, Pz, PO7, Oz, and PO8 on the cortex. Channels were referenced via bilateral mastoid references. For two groups, the EEG streams of two participants were lost and omitted from the analysis. Furthermore, two groups had streams from two participants which terminated prematurely 20 minutes after the start of the task. Four

Table 5: Number of utterances for each higher-order dimension of Marks et al.'s taxonomy of team processes [71] for each group. Utterances were averaged across group members for each category per 30 s window, and summed across the experiment session; their sum per group is shown in the "Total" column. Values marked with an asterisk are statistical outliers based on the IQR method (Q1–Q3 range).

Group	Transition	Action	Interpersonal	General Teamwork	Total
1	131.3	170.5	244.5	123.5	669.8
2	144	132	279	126	681
3	35	45.5	81	21.3	182.8*
4	63.5	80	150	66.5	360
5	183.3	241.3*	488.7*	166.7	1080*
6	98	144.5	165.5	101.8	509.8
7	142	100	219	137.8	598.8
8	90.5	149.5	179.5	90	509.5
9	92.7	144	323.3	118.7	678.7
10	133.3	114	260.3	81.3	589
11	150.7	128.7	138	115.7	533
12	84.9	101.2	189.5	45.3	420.8

groups had one participant whose data had 1-2 excessively noisy channels, which were omitted from the analysis on an individual basis (i.e., participants with clean signals did not have these channels omitted). Except in the case of two participants with shortened 20 min sessions due to equipment malfunctions, the EEG recordings for all participants were approximately 1 hour long when restricted to the duration of the task. In total, we collected $\sim$(40 participants $\times$ 3600 s + 2 participants $\times$ 1200 s) $\times$ 250 Hz = 36,600,000 total data points.

4.4.2 Preprocessed Data. We 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 of padding before and after the task to mitigate any edge effects, then filtered from 0.1–40 Hz (as recommended in [118]), after which excess padding was removed. 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. This filter uses 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 [24]. Note that for some groups, one of the recording computers wrote data more slowly as a result of high usage of system resources, resulting in a loss of data at the end of the task. Data for all participants were truncated to the length of the shortest participant stream.

4.4.3 Derived Individual-Level EEG Measure: Power Spectral Density. 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 [53]. The PSD of the alpha frequency band (8-12 Hz) from the occipital and parietal lobe and the PSD of theta bands (4-7 Hz) 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 (13-30 Hz) power seems to indicate the end of a cognitive task.

PSD values were calculated by computing the mean band power for each of the relevant frequency bands (theta, alpha, and beta) individually over 1-second windows using MNE's [37] `compute_psd` function.

Table 6: Summary statistics for the task engagement index (TEI) per team.

Group	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
1	0.00	0.09	0.15	0.20	0.26	2.18	0.17
2	0.01	0.09	0.14	0.17	0.22	1.30	0.12
3	0.00	0.08	0.14	0.18	0.23	6.88	0.18
4	0.01	0.07	0.12	0.15	0.19	2.21	0.14
5	0.01	0.06	0.11	0.14	0.18	2.40	0.13
6	0.00	0.06	0.10	0.13	0.16	1.30	0.10
7	0.01	0.07	0.11	0.14	0.17	1.06	0.10
8	0.00	0.05	0.09	0.10	0.14	0.90	0.08
9	0.00	0.07	0.13	0.14	0.20	2.64	0.11
10	0.01	0.08	0.13	0.16	0.20	1.36	0.11
11	0.00	0.09	0.17	0.30	0.31	6.38	0.44
12	0.01	0.06	0.16	0.22	0.32	2.14	0.21

4.4.4 Derived Individual-Level EEG Measures of Engagement and Workload. The Task Load Index (TLI) and Task Engagement Index (TEI) are indices derived from these observations. The Task Engagement Index is defined as the ratio of mean beta power to (mean alpha power + mean theta power) (Equation 1). 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 [9, 30, 77, 81].

$$\mathrm{TEI} = \frac{\beta}{\alpha + \theta} \tag{1}$$

The Task Load Index is defined as the ratio of the mean frontal midline theta power to the mean parietal alpha power (Equation 2), and has been shown to increase with cognitive load during a task [36, 46, 56].

$$\mathrm{TLI} = \frac{\theta}{\alpha} \tag{2}$$

Table 7: Summary statistics for the task load index (TLI) per team.

Group	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
1	0.02	1.36	2.70	5.60	5.56	277.96	9.91
2	0.01	1.19	2.54	4.67	5.26	221.72	7.81
3	0.07	1.65	3.28	8.08	6.97	332.01	16.75
4	0.03	0.75	1.67	4.16	3.76	191.09	8.85
5	0.04	2.15	4.62	8.50	10.52	223.40	11.04
6	0.02	1.20	2.63	7.06	6.54	259.85	13.80
7	0.02	1.41	2.83	4.82	5.62	105.72	6.69
8	0.03	1.69	4.04	11.53	12.37	343.44	19.39
9	0.02	1.50	3.95	11.81	12.41	415.92	21.01
10	0.06	1.43	2.67	4.81	5.08	295.04	8.05
11	0.02	0.93	2.07	4.34	4.57	149.03	7.39
12	0.06	1.38	2.88	6.02	6.53	176.97	9.24

As both indices only rely on the PSD of the frontal and parietal lobe, we only considered the Fz, Cz, Oz, PO7, Pz, and PO8 channels for this analysis, leaving 1500 samples per 1-second window (250 samples for 6 channels) and approximately 3600 windows per participant per 1-hour experiment session. The focus on specific frequency bands also reduces the impact of electric noise and outside interference in other bands. Using these band power values, we then calculated the TEI and TLI for each window and individual participant.

Summary statistics for the task engagement index (TEI) and task load index (TLI) individual neural measures are provided for each team in Tables 6 and 7; the relative distributions of each of the neural indices across group members and teams are shown in Figures 4 and 5. Overall, TEI exhibited lower variability than TLI at the group level, but both indices showed variability across individual team members.

4.4.5 Derived Team-Level EEG Measure: Multidimensional Recurrence Quantification Analysis (MdRQA). Multidimensional recurrence quantification analysis (MdRQA) is a technique for measuring the structural and temporal characteristics of multivariate 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. Work 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. [29] 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 [71] team processes framework, we focus on this metric for our analysis as well.

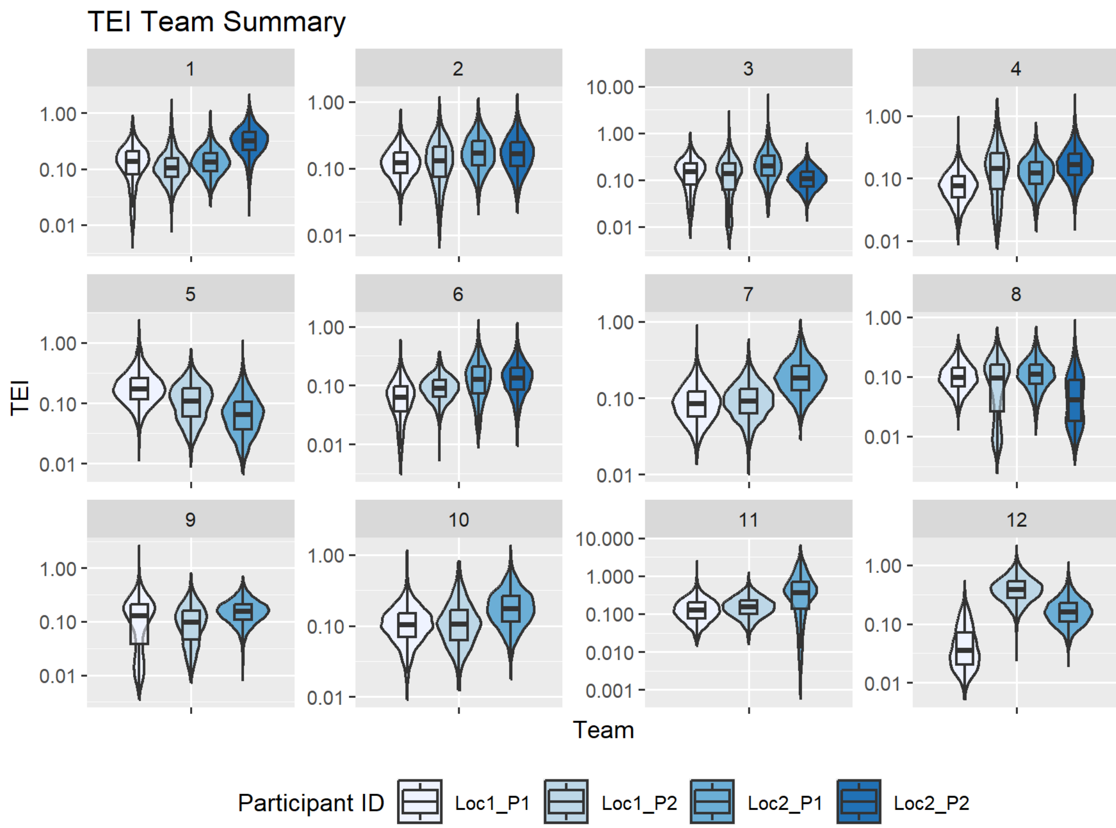

Figure 4: Distribution of Engagement across teams and team members. Each violin shows the Task Engagement Index (TEI) values from a single participant. Participant IDs describe the experiment site (Loc1 - Worcester Polytechnic Institute or Loc2 - University of Bremen) and participant number (P1 or P2) at that site. Each unique participant ID corresponds with a single recording device and EEG headset.

MdRQA entails transforming the time series of interest into a phase space representation via time-delayed embedding. This representation is then used to generate a recurrence plot, a map of all time points that are *recurrent*, or closer than some user-defined distance threshold, or radius. 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.) [117]. The metric used for our analysis, the recurrence rate, is the percentage of recurrent points in the recurrence plot.

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). The time delay and embedding dimension were estimated using the procedure outlined in [114] and the associated MATLAB code[6].

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.

Next, we estimated the optimal radius ϵ, a non-trivial process for our dataset. As recurrences can be meaningful at various distance scales in the phase space, we adopted a data-driven approach used by Yang et al. [122]. 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

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

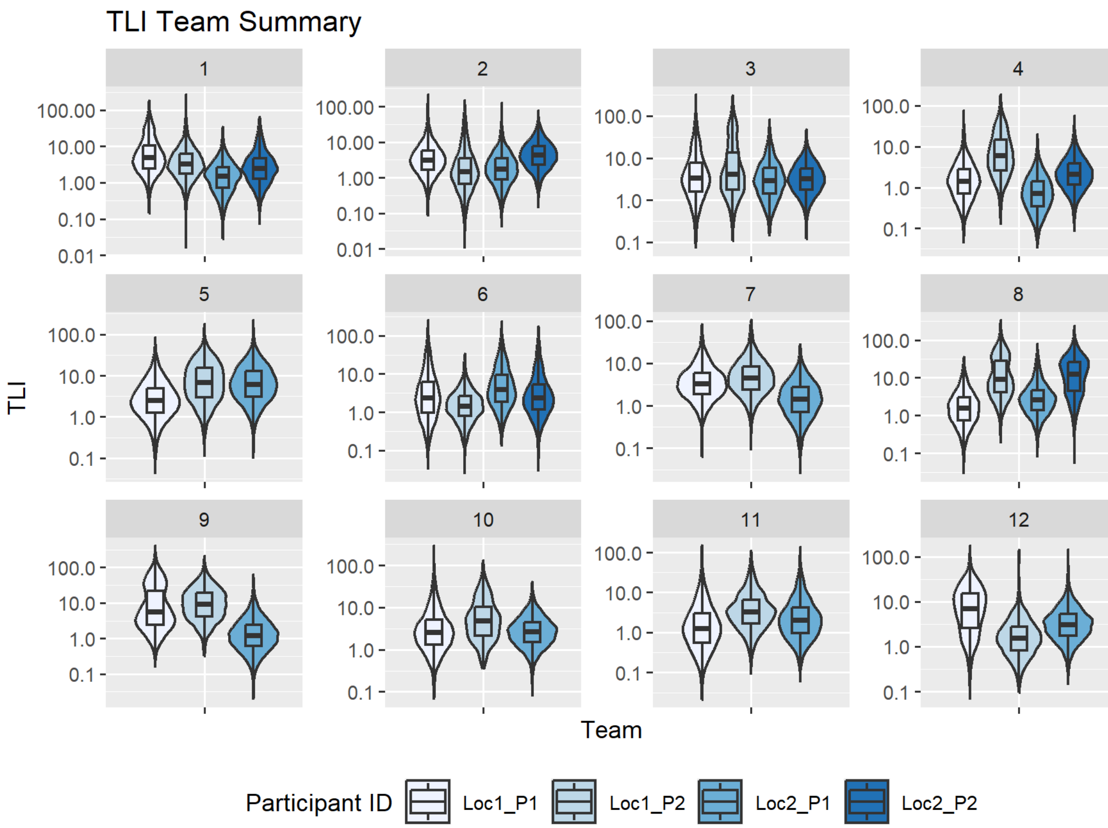

Figure 5: Distribution of Workload across teams and team members. Each violin shows the Task Load Index (TLI) values from a single participant. Participant IDs describe the experiment site (Loc1 - Worcester Polytechnic Institute or Loc2 - University of Bremen) and participant number (P1 or P2) at that site. Each unique participant ID corresponds with a single recording device and EEG headset.

to leverage multiprocessing to more efficiently calculate the recurrence metrics of our large dataset. We first construct a collection of N phase-randomized surrogates, using the amplitude-adjusted Fourier transform method [99] implemented by the `surrog` function in `wsyn`.

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 [122] as our scoring function.

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-second windows (with 15 s overlap) for each group. As in the optimization procedure above, first all channels 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, resulting in a recurrence rate time series for each group.

4.4.6 Derived Team-Level EEG Measure: 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 [106].

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*. We used the second estimator provided by Kraskov et al. [62], 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

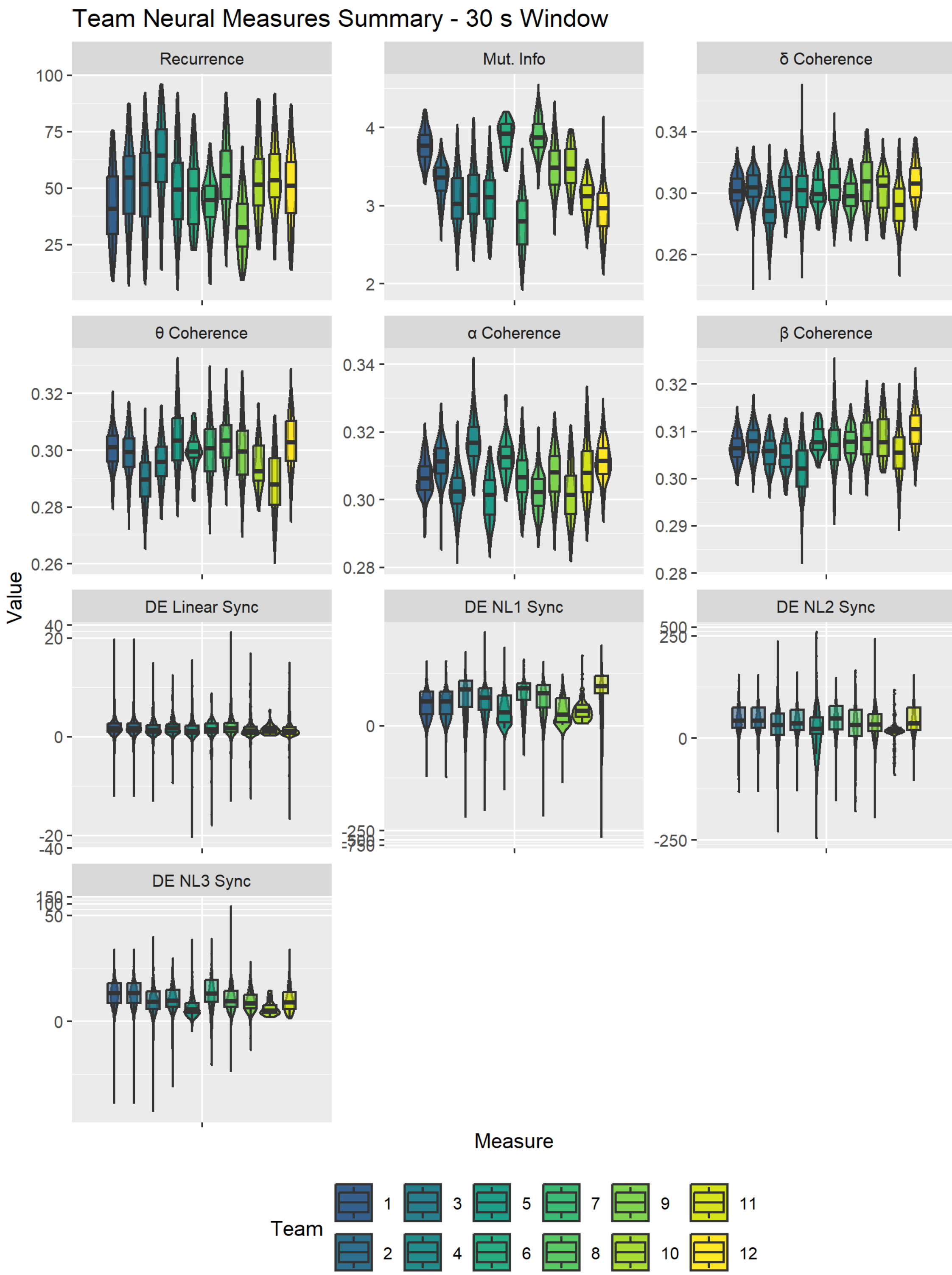

Figure 6: Distribution of values for each team's group-level neural measures. Driver-empath synchrony measures are shown on a symmetric log scale to improve visibility.

Gaussianity, amplitude comparability, or a particular underlying distribution.

We used 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 [34, 102]. We calculated the mutual information for each pair of participants for all eight channels in successive 30-second half-overlapping windows. 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, resulting in a single time series per group.

4.4.7 Derived Team-Level EEG Measures of Joint Processes: 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 for measuring inter-brain synchrony [22, 43], across a wide variety of tasks and contexts [22, 68, 70, 79, 80, 93] using fMRI, EEG, and fNIRS. WTC is especially suitable for assessing synchrony in physiological data because its time resolution varies with 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.

We used Grinsted's [38] `wavelet-coherence` MATLAB toolbox for our analysis as well as Hu & Si's [50] `Multiple Wavelet Coherence` toolbox for 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 30-second half-overlapping windows 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 for all eight channels by averaging the subject-pair coherence values for the relevant channel sets, resulting in four time series (one per frequency band) for each group of participants.

4.4.8 Derived Team-Level EEG Measure: Driver-Empath Model of Teamwork. Guastello & Peressini [42] 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 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.

We adapted the `ProcPlayer`[7] 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 [41] which can be used to calculate the **P** matrix, which contains information about the group's effects on individuals. 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 [42]. The second is a nonlinear model (Nonlinear Model 1; NL1), which 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 [40]. The third is a nonlinear logistic map model (Nonlinear Model 2; NL2), used by [111] 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 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. Finally, the fourth model (Nonlinear Model 3; NL3) is an exponential variant of the logistic map of NL2.

For successive half-overlapping 30-second windows of data, 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, which is typically enough to capture events of interest in biological data [42, 57]. We then calculated the mean of these metrics across all eight channels 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.

4.4.9 Summary of Team-Level Neural Measures. Team-level neural measures experience a considerable degree of variability between teams, with some exhibiting more prominent differences than others. Table 8 contains summary statistics for each of the group neural measures, and their distributions for each team are shown in Figure 6. Delta and beta coherence, as well as DE linear, NL2, and NL3 synchrony, were most consistent across groups. All others varied modestly, with overlapping ranges of observed values in the distribution tails but clear differences in both relative dispersion and their respective interquartile ranges.

5 Linking Neural Signals to Task Performance and Team Processes

Understanding how brain activity relates to team behavior is essential for designing adaptive systems. From our unique dataset, we integrate individual EEG metrics (Task Engagement Index, Task Load Index) with group-level synchrony measures (recurrence rate, mutual information, wavelet transform coherence, and driver–empath

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

Measure	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
Recurrence	4.77	36.92	49.34	49.48	61.87	95.95	17.71
Mut. Info	1.92	2.96	3.28	3.29	3.64	4.55	0.46
δ Coherence	0.237	0.291	0.301	0.300	0.310	0.371	0.015
θ Coherence	0.260	0.291	0.298	0.298	0.305	0.332	0.011
α Coherence	0.281	0.302	0.307	0.307	0.313	0.342	0.008
β Coherence	0.282	0.304	0.307	0.307	0.310	0.325	0.005
DE Linear	-23.32	0.08	0.16	0.24	0.27	26.96	1.37
DE NL1	-441.99	0.36	0.88	1.28	1.79	106.30	11.61
DE NL2	-218.59	0.22	0.45	0.87	0.87	334.84	12.99
DE NL3	-20.75	0.32	0.51	0.66	0.79	85.41	2.33

Table 8: Summary statistics across all groups for group neural measures.

synchrony) and behavioral indicators (task performance and speech-based team process coding). These methods provide a foundation for modeling coordination, workload, and breakdowns in collaboration. While individual-level brain activity highlights how the cognitive state of different people fluctuates and interacts with the group behavior, the team-level brain activity measures quantify connectedness and synchronization (or lack thereof) within the team.

To investigate how our neural measures related to outcomes of interest, and to demonstrate how our dataset can enable the design of future adaptive systems, we conducted a series of exploratory analyses. In Section 5.1, we report results of correlation and regression analyses to determine the relationship between neural measures and team design scores, while in Section 5.2 we report the correlations between neural measures and team processes. Finally Section 5.3, we develop multi-level generalized mixed effects regression models as proofs-of-concept to demonstrate how team and individual neural measures can be used to predict levels of team processes, providing a foundation for real-time adaptive systems supporting collaboration in multi-person teams.

5.1 Relationship Between Neural Indices and Team Performance

Linking neural measures to task performance outcomes could help identify whether cognitive states are predictive of success, leading to adaptive systems that aim to boost team outcomes. To perform correlations comparing each neural measure with each team's design score, we calculated the group average of all time windows across the duration of the experiment as well as across group members in the case of the individual Task Engagement Index (TEI) and Task Load Index (TLI).

5.1.1 Team-Level Neural Measures. We found moderate positive correlations and moderate decision tree goodness of fit between team design scores and DE NL3 synchrony ($r = .59$, $DT\ r^2 = .52$), DE NL1 synchrony ($r = .55$, $DT\ r^2 = .33$), and mutual information ($r = .23$, $DT\ r^2 = .45$). Conversely, we observed moderate negative correlations between DE NL2 synchrony ($r = -.50$, $DT\ r^2 = .18$), δ coherence ($r = -.36$, $DT\ r^2 = .33$), and θ coherence ($r = -.36$, $DT\ r^2 = .19$). These results suggest that several group neural measures can be directly linked to team performance.

5.1.2 Individual-Level Measures. Both individual indices showed small but positive associations with design quality, with Task Load Index ($r = .34$) showing a slightly stronger relationship than Task Engagement Index ($r = .26$). 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 the Decision Tree (DT) Regressor provided by `scikit-learn` [86], generating different models using combinations of the individual neural indices and mean or maximum escape room familiarity as predictors. 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 the data.

Across models, Task Engagement Index and Task Load Index outperformed measures of prior experience (e.g., average or maximum escape room familiarity) in predicting design scores. The strongest individual predictor was TEI, with $r^2 = .13$. The result was unchanged when both indices were used as predictors, as well as when prior experience was included in the model.

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.

5.2 Exploring Relationships Between Neural Measures and Team Processes

We investigated team processes as outcome measures, exploring the extent to which changes in the brain dynamics of group members were related to changes in their behavioral dynamics during collaboration. For example, we were interested in whether the interaction behavior, such as joint attention or interpersonal dialog, could be predicted from the neural data. Examining neural data alongside team processes reveals how collaboration unfolds over time, enabling fine-grained interventions that support teams in critical moments. In the future, this could enable post-hoc meeting analysis or real-time meeting support. We analyzed the correlation between team-level neural measures and team processes after averaging the data across time windows. This approach allowed us to identify any potential overall effects, with several robust effects emerging.

The correlation analyses below were conducted after removing data for outlier groups with unusually high or low communication (Groups 3 and 5; see Table 5), which substantially strengthened several observed relationships. The results of these analyses with outliers included are provided in Supplement S1 in the Supplementary Material.

5.2.1 Overall Team-level Neural Measures and Team Processes. Table 9 contains correlation coefficients from comparing the time-averaged team neural measures and team processes.

Mutual information displayed a strong positive correlation with action processes ($r = .71$), while α coherence exhibited a moderate negative correlation ($r = -.47$). δ coherence was moderately correlated with interpersonal processes ($r = .60$), while DE NL1 synchrony exhibited a strong negative correlation with ($r = -.83$). Recurrence rate was likewise negatively correlated with interpersonal processes ($r = -.71$), as well as with action processes and general teamwork ($r = -.55$ and $r = -.54$, respectively).

While exploratory, these patterns represent promising results, corroborating findings from prior work linking recurrence, mutual information and coherence with collaborative behavior at the group level.

Table 9: Pearson correlation between higher-order team process dimensions and mean team neural measures (outliers removed).

	Transition	Action	Inter-personal	General Team-work
Recurrence	-0.24	-0.55	-0.72	-0.54
Mut. Info	-0.24	0.71	0.05	-0.06
δ Coherence	-0.27	-0.16	0.60	-0.01
θ Coherence	-0.37	0.20	0.22	-0.09
α Coherence	-0.42	-0.48	-0.39	-0.25
β Coherence	-0.14	-0.02	0.31	-0.29
DE Linear	0.11	0.14	0.22	0.11
DE NL1	-0.09	-0.30	-0.83	-0.31
DE NL2	-0.24	-0.02	0.03	0.37
DE NL3	-0.29	0.41	-0.45	0.05

5.2.2 Overall Individual-level Neural Measures and Team Processes. The correlation coefficients from comparing the time-averaged individual neural measures and team processes are shown in Table 10. TEI was negatively correlated with all team process dimensions with r-values ranging from -.23 to -.29, while TLI was positively correlated with all team process dimensions (strongest for interpersonal processes with $r = .35$).

Table 10: Pearson correlation between higher-order team process dimensions and mean individual neural measures (outliers removed).

	Transition	Action	Interpersonal	General Teamwork
TEI	-0.29	-0.27	-0.28	-0.23
TLI	0.08	0.12	0.35	0.12

5.3 Foundations for Adaptive Interfaces that Respond to Team-Level Cognitive and Behavioral States

The physical setup and data acquisition and synchronization pipeline that we presented here demonstrated a proof-of-concept that neural data can be successfully captured for distributed teams, even in different continents. This demonstrates the technical feasibility of brain-based interventions assisting real-world distributed teams. In the analysis described above, we verified that there are interpretable relationships between team-level and individual-level neural measures and final team outcomes (task performance) as well as intermediate dynamics (team processes). Furthermore, building on a streaming middleware like LSL and studying measures on relatively short time windows that do not require compute-intense training or calibration from annotated data (e.g., in contrast to methods based on machine learning), as was the case for all of our measures save recurrence (due to the need to derive an optimal radius and embedding parameters) enable the seamless transition to real-time BCIs.

Even with modest predictive power, this opens up opportunities for lightweight monitoring that could drive adaptive user interfaces that better support teamwork. For example, a system could provide adaptive notifications when detecting early signs of disengagement or overload, or it could drive dynamic task allocation to adjust roles or suggest breaks when workload imbalance is detected. It could also be used to inform AI agents—such as ChatGPT, which was used in the study—to prompt coordination at appropriate times, based on synchrony measures.

In order to demonstrate the potential for our multimodal dataset to facilitate the development of real-time adaptive systems supporting collaboration, we developed multi-level regression models for team processes integrating all of our data as a proof of concept. While more sophisticated techniques such as neural networks or online ensemble methods would be necessary to achieve real-world utility, generalized linear regression models are a useful starting point because they are able to provide transparent insight into feature importance, are efficient to construct and train, and can be easily updated with new data.

The goal is to model individual-level team processes as a function of both individual- and group-level measures, while also incorporating cross-sectional information about participants and accounting for the nested structure of our dataset. At the individual level, data for team processes is *count* data (the mean number of utterances for a given time window); therefore a model suited for such data, such as a Poisson or negative binomial model, is necessary. It is also worth noting that at the individual level our dataset exhibits considerable zero-inflation for team processes (with the percentage of zeros 77%, 73%, and 65% for transition, action, and interpersonal processes, respectively). Zeros occur in the data whenever a participant did not speak during a given time window, and can arise because, e.g., the participant was listening while another team member was speaking, or was engaged productively with the task, possibly while interacting non-verbally with other team members via Miro. While it is possible to omit them from analysis, doing so would result in biased estimates, since the true data distribution is misrepresented, so it is best to account for them in the regression model.

We used the `glmmTMB` package in R to develop our models [15]. We chose the `nbinom2()` negative binomial family, as Poisson models experienced overdispersion (with dispersion ratios of 2.858, 2.273, and 3.253 for transition, action, and interpersonal processes, respectively), and incorporate a zero-inflation component to accurately model zeros. Table 11 lists predictors used in the regression models. We constructed three models, each modeling either transition, action, or interpersonal processes for individual participants as a function of individual- and group-level neural measures and cross-sectional data in the conditional component; the zero-inflation component modeled team-process zeros as a function of all neural measures as well as the number of group members. Additionally, we included random intercepts for both participant ID and group number in both model components to account for individual differences between subjects and teams. All numeric predictors (all neural measures and PMERQ scores) were z-scored prior to model fitting to aid numerical stability and facilitate the interpretation of coefficients by placing them on a common scale.

The full set of regression models and complete model output are provided in Supplement S2 in the Supplementary Material; here we only discuss significant neural predictors of the conditional model for brevity. The log likelihoods for transition, action, and interpersonal process models were -4890.3, -5516.2, and -7230.2, respectively. For transition processes, TEI (Est. = .31, SE = .095, z = 3.281, $p < .01$) and recurrence rate (Est. = .09, SE = .042, z = 2.262, $p < .05$) were positive predictors. TEI (Est. = .27, SE = .088, z = 3.083, $p < .01$) and θ coherence (Est. = .06, SE = .030, z = 1.997, $p < .05$) were positive predictors for action processes, while DE NL1 Synchrony (Est. = -.038, SE = .017, z = -2.175, $p < .05$) was the only significant negative predictor. Finally, TEI (Est. = .14, SE = .068, z = 2.010, $p < .05$) is a positive predictor for transition processes, while recurrence rate (Est. = -.11, SE = .03, z = -3.781, $p < .001$) and mutual information (Est. = -.15, SE = .039, z = -3.972, $p < .001$) were negative predictors.

These findings, combined with the publicly available dataset and infrastructure, provide a foundation for the HCI community to design, validate, and benchmark adaptive collaborative systems that respond to team-level cognitive and behavioral states.

6 Discussion and Limitations

This paper builds a foundation for capturing team cognition and developing brain-based adaptive team interfaces. By developing the experiment paradigm and carefully documenting all data acquisition techniques, as well as pre-processing and analysis methods, this dataset provides a reproducible foundation for engineering adaptive collaborative systems, allowing researchers to benchmark synchrony metrics and prototype real-time interventions without costly multi-site data collection. Further, the dataset can be used as a resource to better understand users during teamwork. We describe our findings, but the dataset contains additional metrics that could be further explored to answer new research questions.

We saw 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. 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 Task Load Index and transition processes as well as overall process-relevant communication. 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.

Mutual information was strongly correlated with action processes, supporting prior work from Stevens & Gallway [106], 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. [29], 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 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. Finally, α coherence's negative association with action processes and δ coherence's positive association with interpersonal processes align with observations by [107] and [90], with α asymmetry indicative of the presence of leader-follower dynamics, and increased δ coherence associated with joint attention, verbal communication, and shared motor activity.

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. While our dataset enabled analysis of inter-brain dynamics, these hyperscanning methods are vulnerable to confounds such as common input effects and EEG-specific issues like volume conduction (Section 2.3). Although we applied standard preprocessing, future work should incorporate additional controls such as surrogate data generation, phase randomization, spatial filtering, and baseline comparisons to better isolate interpersonal coupling [17, 21, 124].

We also see opportunities to connect this work with interactive visual analytics platforms. The BrainEx system [48] introduced scalable, distributed methods for fast exploration of large brain-signal collections, enabling users to cluster signals, perform warped-distance similarity search, and examine patterns across conditions. Extending this class of tools to support multi-person datasets, such as the synchronized EEG, transcripts, and team-process measures presented here, would allow researchers to probe when similar cognitive states or coordination patterns recur across individuals or groups.

While our analysis of team processes is theoretically grounded, 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

Table 11: Predictors used in regression models. Individual cross-sectional variables were omitted from the zero-inflation component. DE NL1 Driver and Empath scores were used for predicting action processes because they yielded the lowest AIC of candidate models using Linear, NL2, or NL3 scores.

Predictor Category	Predictors
Individual neural measures	TEI, TLI, Linear Driver Score and Linear Empath Score (NL1 for action processes)
Individual cross-sectional variables	Age, Gender, Ethnicity, Education, PMERQ Scores for all 10 dimensions, Prior Escape Room Familiarity
Group neural measures	MdRQA Recurrence Rate, Mutual Information, $\delta/\theta/\alpha/\beta$ Coherence, DE Linear/NL1 Synchrony
Group cross-sectional variables	# Group Members

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 explore more information about the non-verbal dimensions of participant interactions, using the dataset described here. One straightforward first step is to categorize participant interactions, while they used Miro to document their designs. Adapting the approach from Aytes [4] 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 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 [100], or we could develop a coding scheme to categorize the types of interactions people had [44]—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 [35].

7 Conclusion

This paper provides a publicly available multimodal dataset capturing EEG and behavioral data from distributed teams engaged in creative collaboration, along with methods and infrastructure for synchronized recording. This resource enables the HCI community to design, validate, and benchmark adaptive collaborative systems that respond to team-level cognitive and behavioral states.

Acknowledgments

We thank the Radcliffe Institute of Advanced Study at Harvard University and Radcliffe Research Partners Itzel Sanchez and Henry Huang. This material is based on work partially supported by the U.S. National Science Foundation (NSF) under Grant No DGE-1922761 and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 459360854 (project name: „Lifespan AI: From Longitudinal Data to Lifespan Inference in Health (DFG FOR 5347)"). We acknowledge the use of generative AI tools (ChatGPT, Gemini, Microsoft 365 Copilot) to assist with editing author-generated text during the preparation of this manuscript.

References

[1] 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. doi:10.1145/2642918.2647414

[2] 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.

[3] 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). doi:10.3389/fnins.2020.00584

[4] 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. doi:10.1007/BF00823363

[5] Lain Bailey. [n. d.]. OBS Studio. https://obsproject.com/

[6] 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

[7] 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. doi:10.1145/1753326.1753511

[8] 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. doi:10.1037/amp0000305

[9] 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.

[10] Luigi Bianchi, Francesco Gambardella, Chiara Liti, and Veronica Piccialli. 2019. Group study via collaborative BCI. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 272–276. doi:10.1109/SMC.2019.8914482 ISSN: 2577-1655.

[11] Alexander Bick, Adam Blandin, and David J Deming. 2024. *The Rapid Adoption of Generative AI*. Working Paper 32966. National Bureau of Economic Research. doi:10.3386/w32966

[12] 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. doi:10.1080/1463922X.2022.2086644

[13] 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. doi:10.1080/2326263X.2015.1100038

[14] Elisabeth Brauner, Margarete Boos, and Michaela Kolbe. 2018. *The Cambridge Handbook of Group Interaction Analysis*. Cambridge University Press. Google-Books-ID: KnNqDwAAQBAJ.

[15] Mollie Brooks, Ben Bolker, Kasper Kristensen, Martin Maechler, Arni Magnusson, Maeve McGillycuddy, Hans Skaug, Anders Nielsen, Casper Berg, Koen van Bentham, Nafis Sadat, Daniel Lüdecke, Russ Lenth, Joseph O'Brien, Charles J. Geyer, Mikael Jagan, Brenton Wiernik, Daniel B. Stouffer, Michael Agronah, Hatice Tül Kübra Akdur, Daniel Sabanés Bové, and Nikolas Krieger. 2025. glmmTMB: Generalized Linear Mixed Models using Template Model Builder. https://cran.r-project.org/web/packages/glmmTMB/

[16] Melanie S. Brucks and Jonathan Levav. 2022. Virtual communication curbs creative idea generation. *Nature* 605, 7908 (May 2022), 108–112. doi:10.1038/s41586-022-04643-y

[17] Adrian Philip Burgess. 2013. On the interpretation of synchronization in EEG hyperscanning studies: a cautionary note. *Frontiers in Human Neuroscience* 7 (Dec. 2013). doi:10.3389/fnhum.2013.00881

[18] 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. doi:10.1145/3059454.3059455

[19] Aaron Chatterji, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. *How People Use ChatGPT*. Working Paper 34255. National Bureau of Economic Research. doi:10.3386/w34255

[20] 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

[21] Artur Czeszumski, Sara Eustergerling, Anne Lang, David Menrath, Michael Gerstenberger, Susanne Schuberth, Felix Schreiber, Zadkiel Zuluaga Rendon, and Peter König. 2020. Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience* 14 (Feb. 2020). doi:10.3389/fnhum.2020.00039

[22] 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. doi:10.1523/ENEURO.0268-21.2022

[23] 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. doi:10.1088/1741-2560/13/4/046022

[24] Arnaud Delorme. 2023. EEG is better left alone. *Scientific Reports* 13, 1 (Feb. 2023), 1–12. doi:10.1038/s41598-023-27528-0

[25] 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. doi:10.1145/2851581.2892329

[26] 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. doi:10.1007/978-3-642-21852-1_35

[27] 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. doi:10.1037/1089-2699.10.4.249

[28] 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. doi:10.1371/journal.pone.0213516

[29] 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. doi:10.1145/3579598

[30] 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. doi:10.1023/A:1009566809021

[31] 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. doi:10.1145/3290605.3300619

[32] 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. doi:10.1016/j.tsc.2016.05.009

[33] Leo Galway, Paul McCullagh, Gaye Lightbody, Chris Brennan, and David Trainor. 2015. The Potential of the Brain-Computer Interface for Learning: A Technology Review. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. 1554–1559. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.234

[34] 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. doi:10.1109/TIT.2018.2807481

[35] 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. doi:10.1016/j.leaqua.2017.11.003

[36] 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. doi:10.1080/14639220210159717

[37] 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. doi:10.3389/fnins.2013.00267

[38] 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. doi:10.5194/npg-11-561-2004

[39] g.tec medical engineering GmbH. 2024. Unicorn Hybrid Black. https://www.gtec.at/product/unicorn-hybrid-black/

[40] Stephen J. Guastello. 2016. Physiological Synchronization in a Vigilance Dual Task. *Nonlinear Dynamics, Psychology, and Life Sciences* 20, 1 (Jan. 2016), 49–80.

[41] 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.

[42] 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).

[43] 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. doi:10.1016/j.neuroimage.2023.120354

[44] 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. doi:10.1145/3613904.3642133

[45] Elwira A. Hałgas, Kyana H. J. van Eijndhoven, Josette M. P. Gevers, Travis J. Wiltshire, Joyce H. D. M. Westerink, and Sonja Rispens. 2023. A Review of Using Wearable Technology to Assess Team Functioning and Performance. *Small Group Research* 54, 1 (Feb. 2023), 41–76. doi:10.1177/10464964221125717

[46] Anu Holm, Kristian Lukander, Jussi Korpela, Mikael Sallinen, and Kiti Müller. 2009. Estimating Brain Load from the EEG. *TheScientificWorldJournal* 9 (02 2009), 639–51. doi:10.1100/tsw.2009.83

[47] 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.

[48] 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. doi:10.1145/3534516

[49] 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. doi:10.1080/21650349.2021.1976678

[50] 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. doi:10.5194/hess-20-3183-2016

[51] 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. doi:10.1109/TAFFC.2019.2901456

[52] 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. doi:10.1016/j.knosys.2013.07.007

[53] 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. doi:10.1371/journal.pone.0242857

[54] 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. doi:10.5465/amp.2019.0040

[55] Robin R. Johnson, Chris Berka, David Waldman, Pierre Balthazard, Nicola Pless, and Thomas Maak. 2013. Neurophysiological Predictors of Team Performance. In *Foundations of Augmented Cognition*, Dylan D. Schmorrow and Cali M. Fidopiastis (Eds.). Springer, Berlin, Heidelberg, 153–161. doi:10.1007/978-3-642-39454-6_16

[56] 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. doi:10.1177/0018720814526617 PMID: 25277022.

[57] 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. doi:10.1177/0018720819874160

[58] 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. doi:10.1145/2998181.2998185

[59] 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. doi:10.1177/10596011241278556

[60] 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. doi:10.1101/2024.02.13.580071

[61] Steve W.J. Kozlowski and Daniel R. Ilgen. 2006. Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest* 7, 3 (Dec. 2006), 77–124. doi:10.1111/j.1529-1006.2006.00030.x

[62] Alexander Kraskov. 2004. Estimating mutual information. *Physical Review E* 69, 6 (2004). doi:10.1103/PhysRevE.69.066138

[63] 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. doi:10.1145/3474696

[64] 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. doi:10.1207/S15326934CRJ1334_06

[65] Octavia Leahy, Emily Kontaris, Natalie Gunasekara, Joy Hirsch, and Ilias Tachtsidis. 2025. Environmental effects on inter-brain coupling: a systematic review. *Frontiers in Human Neuroscience* 19 (July 2025). doi:10.3389/fnhum.2025.1627457

[66] Sunghan Lee, Hohyun Cho, Kiseong Kim, and Sung Chan Jun. 2019. Simultaneous EEG Acquisition System for Multiple Users: Development and Related Issues. *Sensors* 19, 20 (Jan. 2019), 4592. doi:10.3390/s19204592

[67] 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. doi:10.1109/TBCAS.2014.2316224

[68] 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

[69] 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. doi:10.1016/j.neuropsychologia.2019.01.004

[70] 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. doi:10.1002/brb3.2270

[71] 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. doi:10.2307/259182

[72] 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. doi:10.1037/apl0000856

[73] 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. doi:10.5465/amj.2005.18803928

[74] Oded Mayo and Ilanit Gordon. 2020. In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology* 57, 6 (2020), e13574. doi:10.1111/psyp.13574

[75] 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. doi:10.2307/256727

[76] Joseph Edward McGrath. 1984. *Groups : interaction and performance.* Prentice-Hall, Inc., Englewood Cliffs, NJ.

[77] Timothy McMahan, Ian Parberry, and Thomas D. Parsons. 2015. Evaluating Player Task Engagement and Arousal Using Electroencephalography. *Procedia Manufacturing* 3 (2015), 2303–2310. doi:10.1016/j.promfg.2015.07.376 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

[78] 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. doi:10.1016/j.ijhcs.2015.05.002

[79] 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. doi:10.1016/j.cortex.2019.11.020

[80] 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. doi:10.1016/j.neuroimage.2016.03.059

[81] 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. doi:10.1109/COGSIMA.2017.7929581

[82] 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. doi:10.1016/S0013-4694(97)00106-5

[83] 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. doi:10.1177/10731911221134601

[84] 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. doi:10.1016/B978-0-12-374714-3.00014-8

[85] 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. doi:10.1007/978-3-030-77198-0_3

[86] 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

[87] 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.

[88] 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. doi:10.1145/2663204.2663279

[89] Adarsh Pyarelal, Eric Duong, Caleb Shibu, Paulo Soares, Savannah Boyd, Payal Khosla, Valeria A. Pfeifer, Diheng Zhang, Eric Andrews, and Rick Champlin. 2023. The tomcat dataset. *Advances in Neural Information Processing Systems* 36 (2023), 40872–40923. https://proceedings.neurips.cc/paper_files/paper/2023/hash/803d8d4b4a549d0d062fc704f8659ce3-Abstract-Datasets_and_Benchmarks.html

[90] Alejandro Pérez, Manuel Carreiras, and Jon Andoni Duñabeitia. 2017. Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports* 7, 1 (June 2017), 4190. doi:10.1038/s41598-017-04464-4

[91] Yinuo Qin, Richard T. Lee, Weijia Zhang, Xiaoxiao Sun, and Paul Sajda. 2025. Physiologically informed predictability of a teammate's future actions forecasts team performance. *iScience* 28, 5 (May 2025). doi:10.1016/j.isci.2025.112429

[92] Mauricio A. Ramírez-Moreno, Jesús G. Cruz-Garza, Akanksha Acharya, Girija Chatufale, Woody Witt, Dan Gelok, Guillermo Reza, and José L. Contreras-Vidal. 2023. Brain-to-brain communication during musical improvisation: a performance case study. doi:10.12688/f1000research.123515.4

[93] 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.

[94] 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 (Jan. 2021), 43–57. doi:10.1093/scan/nsaa135

[95] 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. doi:10.1073/pnas.2005737118

[96] 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

[97] Lena Ivannova Ruiz-Rojas, Luis Salvador-Ullauri, and Patricia Acosta-Vargas. 2024. Collaborative Working and Critical Thinking: Adoption of Generative Artificial Intelligence Tools in Higher Education. *Sustainability* 16, 13 (2024). doi:10.3390/su16135367

[98] Eduardo Salas, Denise L. Reyes, and Susan H. McDaniel. 2018. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* 73, 4 (2018), 593–600. doi:10.1037/amp0000334

[99] Thomas Schreiber and Andreas Schmitz. 2000. Surrogate time series. *Physica D: Nonlinear Phenomena* 142, 3 (Aug. 2000), 346–382. doi:10.1016/S0167-2789(00)00043-9

[100] 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. doi:10.1007/s00146-021-01335-5

[101] 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. doi:10.1145/2984511.2984578

[102] 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. doi:10.1080/01966324.2003.10737616

[103] 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. doi:10.1145/2207676.2208372

[104] 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. doi:10.1561/1100000078

[105] 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. doi:10.1007/978-3-319-64274-1_8

[106] 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

[107] Caroline Szymanski, Ana Pesquita, Allison A. Brennan, Dionysios Perdikis, James T. Enns, Timothy R. Brick, Viktor Müller, and Ulman Lindenberger. 2017. Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *NeuroImage* 152 (May 2017), 425–436. doi:10.1016/j.neuroimage.2017.03.013

[108] 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. doi:10.1037/amp0000310

[109] 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. doi:10.1145/2160125.2160144

[110] Davide Valeriani, Riccardo Poli, and Caterina Cinel. 2017. Enhancement of Group Perception via a Collaborative Brain–Computer Interface. *IEEE Transactions on Biomedical Engineering* 64, 6 (June 2017), 1238–1248. doi:10.1109/TBME.2016.2598875

[111] 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. doi:10.1075/is.6.1.04val

[112] 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. doi:10.1007/978-1-84882-854-4_9

[113] David A. Waldman, M. K. Ward, and William J. Becker. 2017. Neuroscience in Organizational Behavior. *Annual Review of Organizational Psychology and Organizational Behavior* 4, Volume 4, 2017 (March 2017), 425–444. doi:10.1146/annurev-orgpsych-032516-113316

[114] 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). doi:10.3389/fpsyg.2018.01679

[115] Kai Wang and Jeffrey V. Nickerson. 2017. A literature review on individual creativity support systems. *Computers in Human Behavior* 74 (Sept. 2017), 139–151. doi:10.1016/j.chb.2017.04.035

[116] Yijun Wang and Tzyy-Ping Jung. 2011. A Collaborative Brain-Computer Interface for Improving Human Performance. *PLOS ONE* 6, 5 (May 2011), e20422. doi:10.1371/journal.pone.0020422

[117] Charles Webber and Joseph Zbilut. 2005. Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences* (Jan. 2005).

[118] 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. doi:10.1016/j.jneumeth.2014.08.002

[119] Jessica L. Wildman, Eduardo Salas, and Charles P. R. Scott. 2014. Measuring Cognition in Teams: A Cross-Domain Review. *Human Factors* 56, 5 (Aug. 2014), 911–941. doi:10.1177/0018720813515907

[120] 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. doi:10.1177/1046496420986620

[121] 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. doi:10.1016/j.neuroimage.2018.02.007

[122] 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. doi:10.1016/j.jsv.2015.03.046

[123] 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. doi:10.1145/3544549.3585701

[124] Anna Zamm, Janeen D Loehr, Cordula Vesper, Ivana Konvalinka, Simon L Kappel, Ole A Heggli, Peter Vuust, and Peter E Keller. 2024. A practical guide to EEG hyperscanning in joint action research: from motivation to implementation. *Social Cognitive and Affective Neuroscience* 19, 1 (July 2024), nsae026. doi:10.1093/scan/nsae026

[125] 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. doi:10.1088/1741-2560/8/2/025005

Appendices

A Escape Room Design Rubric

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 12: Scoring rubric used for escape room designs, based on the taxonomy of digital escape rooms by Krekhov et al. [63]. A maximum of ten points could be earned for each of the first eight criteria, and a maximum of four points could be earned for each of the last four, for a total possible score of 100 points.

B Team Collaboration Task Demographics

Table 13: 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 3.1; 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.

ID	Age Range	Gender	Ethnicity	Education	Task 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
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

Table 14: PMERQ scores [83]. Engagement/disengagement focus scores are presented across the five families of emotion regulation strategies from the process model of emotion regulation. Scores for each dimension range from 1 (Lowest agreement) – 6 (Highest agreement). Missing responses from three participants are omitted.

ID	Group	PMERQ Emotion Regulation Strategy Families							
		Situation Selection		Situation Modification		Attentional Deployment		Cognitive Change	
		Engagement-Focused	Disengagement-Focused	Engagement-Focused	Disengagement-Focused	Engagement-Focused	Disengagement-Focused	Engagement-Focused	Disengagement-Focused
		Confront Unpleasant Situations	Avoid Unpleasant Situations	Resolve Conflicts	Avoid Conflicts	Focus Elsewhere	Cognitively Distract	Consider Benefits	Reduce Importance
P1	1	3	4.17	3.83	3.4	2.5	4	4.67	5
P2	1	5	3.33	4.67	2.6	3	2.6	5	2.33
P3	1	4	2.83	4.33	2.8	4	3	4	4.33
P4	1	4	2.33	4.17	2.4	2.75	2.2	4.17	2.67
P5	2	3.5	4	4	2.8	3.5	3.8	3.67	4
P6	2	3.5	4.33	3	4.2	3.5	5.2	3.17	3.33
P7	2	—	—	—	—	—	—	—	—
P8	2	3.75	2.33	3.5	2.6	3.5	4	3.67	3.67
P9	3	4	3.5	4.67	3.8	4.25	3.6	4.5	4.33
P10	3	4.75	4.67	3.83	5	5.5	5.4	5.33	5.67
P11	3	2.5	4.33	3.83	3.8	3.25	3.8	3.17	4
P12	3	1.75	6	3.5	5.4	5	6	3.83	5.67
P13	4	3.25	4.83	4.67	3.8	5	3.2	5.83	5.67
P14	4	2.75	4.83	3.5	4	5.25	4.4	5.67	5
P15	4	4.25	4.33	5	3.6	3	4.6	3.17	4
P16	4	4	4	4.33	3.2	2.5	3.6	4.17	4
P17	5	4	4.67	3.5	3.4	2.75	3.6	4.17	3.33
P18	5	3	3.5	5	2.2	2.5	5.2	4.17	5.67
P19	5	5	5	5	4.6	5	5	5	5
P20	6	5	3.5	4.33	4	4.5	4.8	4.67	4
P21	6	5.5	4.83	4.5	4.8	5	4.4	4.67	5
P22	6	3.75	4.67	4.5	4.2	4.5	4	4.5	5
P23	6	4.5	4	4.33	4.2	3.75	4	4.33	4
P24	7	2.75	5.33	3.67	4.2	3.25	3.6	3.5	5
P25	7	2.5	4.83	4.67	4.4	4.75	4.6	3.5	4.33
P26	7	2.75	5.17	3.83	4	3	4	4.33	4.33
P27	7	5	4.17	4.67	4.8	4.75	5.2	5	5
P28	8	2.25	5.83	3.33	3.4	1.25	5.2	2.5	4.33
P29	8	3.5	4	4.5	4.2	3.75	4.6	4	3.67
P30	8	5	2.33	5	3.2	4	2.4	5.17	3.67
P31	8	2.25	5.17	4.83	5.4	3.25	3.2	5	5
P32	9	3.25	3.33	4.5	4.2	4	4.4	4.5	4.33
P33	9	2	4.5	2.67	4.8	3.25	4.6	4.33	5.33
P34	9	4.5	4	5	4.8	4.5	2.8	5	4
P35	10	3	2.33	4.33	3.4	3.25	3.4	5.17	4.33
P36	10	4.75	2.67	4.17	3.2	3.25	3.2	3.83	3.33
P37	10	3.25	4.17	4.5	3.2	4.25	4.8	4.5	4
P38	11	4	2.83	4.67	1.6	4.25	1.8	5.33	2.67
P39	11	—	—	—	—	—	—	—	—
P40	11	5.25	5	4.33	3	3.75	4.6	4.5	5
P41	12	3.25	4.33	3.17	3.6	3.75	5	3.33	4.33
P42	12	2.5	5.67	3.83	3.6	3.75	4.8		
P43	12	1.25	5.5	4.17	4.4	3.5	5.2	1.5	4.33
P44	12	—	—	—	—	—	—	—	—

Table 15: PMERQ scores [83] (cont). Engagement/disengagement focus scores are presented across the five families of emotion regulation strategies from the process model of emotion regulation. Scores for each dimension range from 1 (Lowest agreement) – 6 (Highest agreement). Missing responses from three participants are omitted.

ID	Group	PMERQ Emotion Regulation Strategy Families — Response Modulation — Engagement-Focused: Support by Emotion Sharing	PMERQ Emotion Regulation Strategy Families — Response Modulation — Disengagement-Focused: Expressive Suppression
P1	1	4.67	4
P2	1	2.67	3.33
P3	1	4.33	2.67
P4	1	3	3.67
P5	2	4.67	3
P6	2	6	3.67
P7	2	—	—
P8	2	3	4
P9	3	4.33	5
P10	3	5.33	4.33
P11	3	2	3.67
P12	3	4.33	4.67
P13	4	5	1.67
P14	4	4	4.67
P15	4	3.33	3.33
P16	4	3.33	3.33
P17	5	3.33	4.33
P18	5	1.333	5.33
P19	5	5	4
P20	6	5.33	3.67
P21	6	4.33	4.67
P22	6	5.33	3
P23	6	4.67	3.67
P24	7	3.67	4.33
P25	7	3.67	4.33
P26	7	6	3.67
P27	7	5	5
P28	8	4.67	3.67
P29	8	3	4.67
P30	8	3.67	5.33
P31	8	2	4.33
P32	9	2.67	5.67
P33	9	4.33	3.67
P34	9	6	1.67
P35	10	5	3.33
P36	10	4.33	4.33
P37	4.67	4.67	
P38	11	4	2.67
P39	11	—	—
P40	11	4	5
P41	12	3	4
P42	12	2	5
P43	12	2.67	2.67
P44	12	—	—