Improving Mission Performance and Readiness for Rapidly Composed Military Teams

Jonathan Sussman-Fort, Joseph Cohn, Jeremiah Folsom-Kovarik, Jeffrey Craighead, Angela Woods, Daniel Wilson, Stephen Kline, Tanner Hilsabeck, Nick Petroff, Dylan Schmorrow
Soar Technology, LLC Orlando, FL
{jsussmanfort, joseph.cohn, jeremiah, jcraighead, angela.woods, daniel.wilson, stephen.kline, tanner.hilsabeck nick.petroff, dylan.schmorrow}@soartech.com

Eduardo Salas
Lila Berger
Maha Khalid
Rice University, Houston, TX
{es32, lb63, mk127}@rice.edu

Chris Berka
Ella Thunen
Advanced Brain Monitoring Carlsbad, CA
{chris, ejthunen}@b-alert.com

Silke Dodel
Deep Science, LLC Boca Raton, FL
silked@posteo.net

ABSTRACT

Measuring and predicting team operational readiness is paramount to establishing effective teams across the military. Despite there being a standard sequence of training for individuals and teams leading to the actual mission execution in operational environments, measuring and understanding performance at the team level to determine readiness remains a challenge. This challenge is even greater for rapidly composed teams with limited experience working together.

The goal of this study is to determine underlying objective biomarkers of high-performing teams. We present our initial results based on 1) data collection from a five-member Marine Fire Support Team (FiST) engaged in a virtual simulation-based training environment; 2) data cleaning and extraction of derived neural measures during key team behaviors; and 3) initial feature development as part of building a causal model to predict readiness for rapidly composed military teams.

Our goal is to enhance team performance prediction by providing a scalable and interpretable framework for assessing team readiness in dynamic operational environments under high-stakes conditions. This approach is grounded in a theoretic framework based on operationalizing *Team Behavioral Competency Theory (TBCT)*, which has been empirically tested across military, aviation, and healthcare settings. This approach maps specific time windows of team behavior to team-level competencies such as Leadership, Mutual Performance Monitoring, Back-up behavior, Team Orientation, and Adaptability and examines underlying neural, behavioral and peripheral physiologic measures. Specifically, this includes the application of novel high-resolution tripolar concentric ring electrode electroencephalogram (tEEG) combined with standard electroencephalogram (EEG), peripheral physiologic and behavioral measurements. Characterization of these multimodal signals is accomplished using derived metrics to identify a set of features to populate the edges of a graphical causal model that predicts overall team performance and readiness at the mission and subtask level. We show that changes in functional connectivity metrics across the team between *pre-stimulus* and targeted *behavioral windows* correlate with a higher subjective mission rating, predominantly in the delta-band suggesting these are key features of team synchrony that could be used to build out a causal model for predicting team performance.

ABOUT THE AUTHORS

Dr. Jonathan Sussman-Fort is a Lead Scientist at SoarTech whose research focuses on human performance and readiness and the use of combined peripheral physiologic and neuroimaging methods to better understand how teams and experts obtain task proficiency. He has led and contributed to research initiatives related to RF communications and low-probability of detection waveforms, the application of reinforcement learning to dynamic spectrum scanning, auditory scene analysis, underwater acoustics simulations and situational awareness for undersea warfare, and Brain-computer interface (BCI) design.

Dr. Joseph Cohn is a retired Navy Captain and Aerospace Experimental Psychologist and current Vice President of Readiness and Medical Solutions at SoarTech and has developed and led numerous research programs focused on enhancing warfighter team and coalition-level performance and readiness in complex battle scenarios. He is an innovator in creating, fostering and guiding diverse multi-disciplinary teams across the DoD to dramatically accelerate the transition of novel human performance-enabling concepts to practical use.

Dr. Jeremiah Folsom-Kovarik is a Senior Scientist at SoarTech and an expert in adaptive AI, user modeling, and cognitive diagnostics, with a PhD in Computer Science. His background in causal inference and multi-paradigm learning—including novelty search and active learning.

Dr. Jeffrey Craighead is a Lead Scientist at SoarTech where his research is focused on applications of AR and VR for the warfighter, stealth assessment systems, AI & machine learning, knowledge representation, autonomous systems, and wearable sensors. He is currently leading the development of multiple DoD-funded mixed reality and simulation products.

Angela J. Woods is a Software Architect at SoarTech with over 25 years of software engineering experience with emphasis on designing system architectures for real-time uses including intelligent training, agent-based simulation, data analytics and game development. She is an expert in dynamic adaptation techniques and her recent work has included using machine learning techniques to build causal models that can augment adaptive intelligent training systems, prototyping novel interactive visualizations of student performance that surface insights for instructors, and quantifying and visualizing human decision-making across a multi-attribute space for both machine-readable and human-interpretable comparisons between human and algorithmic decision-making outcomes.

Stephen Kline is a Software Engineer at SoarTech with experience leading and contributing to research projects involving human-machine interactions, autonomous vehicle behavioral agents, command and control defense systems, and simulation-based training tools.

Daniel Wilson is a Data Science Engineer at SoarTech and develops data-driven solutions to support research in training systems and medical readiness. Daniel specializes in machine learning, statistical modeling, data processing, and applied analytics, with a focus on building interpretable and reliable systems that bridge research and real-world applications.

Tanner Hilsabeck is a Software Engineer at SoarTech and focuses on the preparation and processing of research data for use in AI/ML pipelines. His work has been applied to domains such as human social dynamics, autonomous vehicle agents, and training simulations.

Nick Petroff is a Program Manager at SoarTech with extensive experience across multiple DoD programs and teaming and interaction studies, including agile program management and implementation.

Dr. Dylan Schmorrow is President of Soar Technology and a retired U.S. Navy Captain, recognized for pioneering advances in intelligent autonomous systems, human–machine teaming, and applied AI across training, simulation, and operational domains. He has held senior leadership roles at DARPA, ONR, NAVAIR, and OSD, where he helped transition numerous advanced research efforts into fielded defense capabilities. His current work focuses on integrating AI, modeling and simulation, and decision-support frameworks to accelerate the transition of novel technologies into mission-ready systems.

Maha Khalid is a PhD student in the Human Factors and Human-Computer Interaction program at Rice University working under the supervision of Dr. Eduardo Salas, where her areas of research include human-autonomy teaming, team physiological dynamics, and team performance measurement and assessment. Prior to graduate school, she served as the associate director for the American Psychological Association's Center for Psychology in Schools and Education.

Lila Berger is a PhD candidate at Rice University in the Human Factors and Human-Computer Interaction program. She is advised by Dr. Eduardo Salas in the Making Effective Teams Laboratory, where she is focused primarily on studying teamwork in extreme environments, human error, and human-AI teaming.

Ella Thunen is the Clinical Research Coordinator at Advanced Brain Monitoring. She has experience collecting EEG clinical trial data and working on projects related to neurodegenerative disorders, traumatic brain injuries and psychiatric disorders. She has extensive experience collecting EEG data in clinical trials.

Dr. Eduardo Salas at Rice University provides over 40 years of experience on foundational organizational related team behavior and performance. He has co-authored over 650 journal articles and book chapters and has co-edited 36 books plus authored two books on team training and the science of teamwork.

Chris Berka has extensive experience in measuring team neurodynamics across multiple team sizes in ecologically valid settings. As CEO of ABM she has over 100 peer-reviewed papers and publications, 20+ international patents, a proven track record with over 50 clinical trials and studies, and an extensive normative database.

Dr. Silke Dodel has wide-ranging experience in the application of non-linear dynamical analysis to team-level data streams. She pioneered graph-theoretical functional connectivity analyses and developed objective measures of team performance based on principles from non-linear dynamics and high-dimensional phase space analytics.

Improving Mission Performance and Readiness for Rapidly Composed Military Teams

Jonathan Sussman-Fort, Joseph Cohn, Jeremiah Folsom-Kovarik, Jeffrey Craighead, Angela Woods, Daniel Wilson, Stephen Kline, Tanner Hilsabeck, Nick Petroff, Dylan Schmorrow
Soar Technology, LLC Orlando, FL
{jsussmanfort, joseph.cohn, jeremiah, jcraighead, angela.woods, daniel.wilson, stephen.kline, tanner.hilsabeck nick.petroff, dylan.schmorrow}@soartech.com

Eduardo Salas
Lila Berger
Maha Khalid
Rice University, Houston, TX
{es32, lb63, mk127}@rice.edu

Chris Berka
Ella Thunen
Advanced Brain Monitoring Carlsbad, CA
{chris, ejthunen}@b-alert.com

Silke Dodel
Deep Science, LLC Boca Raton, FL
silked@posteo.net

INTRODUCTION

Problem and Motivation

Measuring and predicting team expert performance is paramount to determining team operational readiness across the military. Despite there being a standard sequence of training for individuals and teams leading to mission execution in operational environments, measuring and understanding performance at the team level to determine readiness remains a challenge. The current state-of-the-art (SoA) in general teaming research includes single modality approaches that focus either on team behavior or team neurodynamics, while SoA in Department of Defense (DoD) team training focuses on subjective measures and objective outcomes. That said, behavior-based approaches rely heavily on qualitative measures, derived from studies conducted in highly controlled settings often with limited ecological validity and of limited generalizable predictive value. Team neurodynamics approaches default to combining measures for *individual* neurophysiological state representation e.g., cognitive load, rely on a single modality e.g., neural only (Pan et al., 2020; Reinero et al., 2021; Sciaraffa et al., 2017; Sussman et al., 2014) and apply post-hoc correlation-based analyses (Antonenko et al., 2019; Stone et al., 2014) or black box Machine Learning (ML) techniques (R. H. Stevens & Galloway, 2016) to examine synchrony in physiological measures during team coordination activities like visual-motor training, (Barraza et al., 2020; Hoehl et al., 2021) student teaming; (Zhang et al., 2021) tutor/tutee interactions, (Stikic et al., 2013). musical performances and social dialogues, (Hu et al., 2018) and flight simulators and submarine simulations (Astolfi et al., 2012; R. Stevens et al., 2010, 2011, 2013). Assessing these signals during team training and performance has proved difficult, and combining these signals, captured from individual team members, into a team level signature that also captures team behavior measures remains a significant challenge. Both behavioral and neurodynamics approaches have provided initial insights into what constitutes effective teams. However, to establish a definitive, generalizable signature that accurately predicts team performance across multiple domains, it is necessary to capture and integrate measurement modalities. Two key challenges include *identifying* the set of behavioral, neural, and peripheral physiologic signals that can form a multimodal signature of team performance and *bridging the gap* between supporting theory on team behavioral competencies and the underlying brain network level interactions that exist between team members in high-performing expert teams. Overcoming these two challenges would allow for the prediction of team level performance across different tasks and sub-tasks.

In this paper, we focus on applying Team Behavioral Competency Theory (TBCT) as a framework for establishing explainable AI tools to develop a generalizable causal predictive model that validates objective neural signatures of high performing teams to address the key challenges stated above. The focus is to supplement existing subjective measures used by instructors with underlying objective biomarkers of team performance with the goal of improving Department of Defense (DoD) team operational readiness to ensure high performance in field operations. Specifically, we present our initial results based on 1) data collection in a relevant DoD team training environment; 2) data cleaning and extraction of derived neural measures during key team actions; and 3) feature development to test the ability to develop a causal model grounded in TBCT focusing on specific neural features that drive teaming processes. We pose two hypotheses to validate our feature development approach each examining the relationship between a specific functional connectivity metric and a subjective team evaluation.

METHODS

Operationalizing 'Team Behavioral Competency Theory'

One of the most refined team performance theories, "team behavior competency theory" (TBCT) developed by Dr. Eduardo Salas, identifies the "Big Five" most critical team competencies - *team leadership*, *mutual performance monitoring, back-up behaviors, adaptability*, and *team orientation* (Salas et al., 2005, p. 5). To provide a theoretical foundation for predicting team level performance across different tasks and sub-tasks we begin with the framework of TBCT. TBCT is based on fundamental cognitive processes like attention, perception, decision-making, and problem-solving and has been used to guide DoD and non DoD team-training capabilities (Zajac et al., 2022). Together these competencies define the mechanisms that contribute to effective team performance.

To guide our approach, we used this scaffolding to map critical tasks within a DoD team setting, specifically Marine Fire Support Teams (FiST) to the more generalized TBCT. A Fire Support Team (FST) is a specialized unit within the military responsible for coordinating and directing various fire support assets, including artillery, close air support, and other indirect fire systems, to support ground operations. They act as a bridge between the ground commander and the supporting arms, ensuring fires are accurately and effectively employed to achieve tactical objectives. We started by mapping FiST specific subtasks e.g., preparation and fire plan development, to each team behavioral competency using a detailed critical task analysis provided by the government test and evaluation team from Expeditionary Cognitive Science Group, Warfighter Performance Department at the Naval Health Research Center (Figure 1). By applying a theory-driven approach we are able to target specific behavioral windows corresponding to FiST subtasks for feature identification to hone our identification process for underlying biomarkers of team performance.

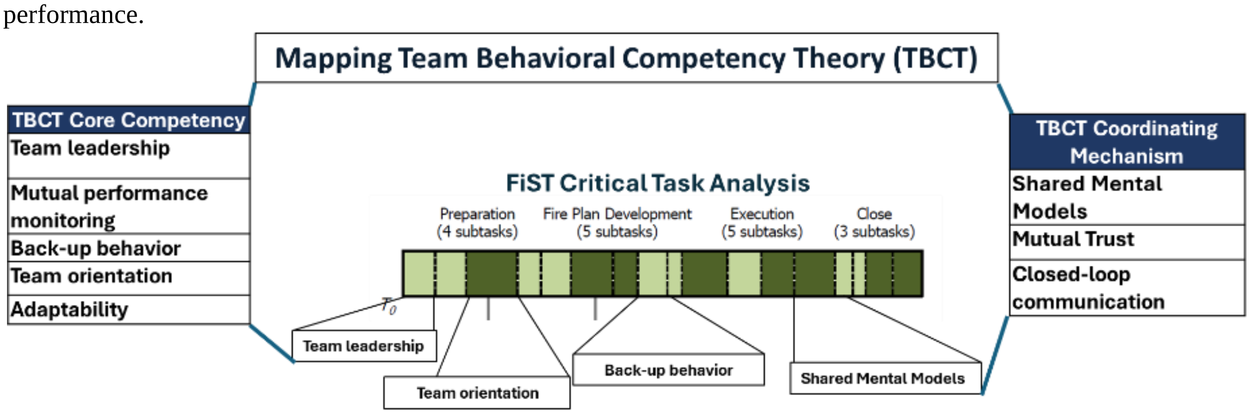

Figure 1: Mapping underlying FiST specific tasks to generalized team theory (TBCT) and specific team competencies and coordinating mechanisms

Identifying and Characterizing Teaming Signals

Team competencies have been theoretically and empirically linked to team performance at the behavioral level, while neural activity in specific brain regions, and peripheral physiologic measures have been shown to correlate to team performance (Berka & Stikic, 2017; Salas et al., 2005). Yet, integrating these behavioral, neural, and biologic signals into a single causally explainable signature of team performance based on an empirically proven theory of team competency remains elusive. As described earlier, a key challenge arises in *identifying* the set of behavioral, neural, and peripheral physiologic signals that can form a multimodal signature of team performance. To address this challenge, we measured 12 total signals as inputs to drive the discovery of a behavioral-neural-peripheral signature (**Table 1**). Neural measures were chosen to examine brain network-level synchronization at multiple levels across all subsets of team members and to understand the cognitive state of individual team members (engagement, workload, drowsiness). Peripheral physiologic measures were chosen to examine team workload, stress, and arousal, while behavioral measures were used to examine verbal and non-verbal interaction underlying team competencies.

Table 1: Multimodal signals measured during FiST Task

Class of Signal	Sensor	Measures/Analysis	Data Collection	Cognitive State
Neural	**Standard EEG 20-channels (B-alert X-24) (1)** **High-resolution tCRE EEG (2)**	Power spectra (1-59Hz)/ channel (delta, theta, alpha, beta, gamma bands) Coherence and Phase-locking within and across brain regions and individual and team members/ channel & band	Software engine compatible with any physio/sensor source or system providing real-time and post-hoc metric output of synched data streams. Min 1 second overlapping windows	Engagement, workload, drowsiness Team synchrony measured across brains using either PSD bands or engagement/workload
Peripheral Physiologic	**ECG (B-alert 2-lead)**	Heart Rate (HR) (3), HR Variability (4); low & high frequency power spectra	HR/HRV 1 second High/low PSD requires 5mins using moving average	Stress, arousal parasympathetic/ Sympathetic balance
	Eye-tracking (Tobii Nano)	Eye Gaze (5), fixations (6), saccades (7), eyeblinks (8), pupil dilation(9), scan path (10)	1-5 seconds depending on metric	Eye gaze directional map, attention, arousal emotional intensity
Behavioral	**Audio/Video**	Speech (10, gesture (11) posture (12)	Automatic speech recognition, gesture recognition	Closed-loop communication, attentiveness

Participants

Participants included four Marine Corps Fire Support Teams (FiST) from the Marine Ground Air Combat Center (MGACC) at Twentynine Palms, CA. Each team consisted of five team members including combined officers and enlisted personnel. Specific roles of the five-member FiST included: FiST Lead, Joint Terminal Attack Controller (JTAC), Fire Support Officer (FSO), Forward Observer (FO) Artillery, and Forward Observer (FO) Mortars.

Testing Environment / JTAC Virtual Trainer (JVT)

Initial testing took place at the Battle Simulation Center at the Marine Ground Air Combat Center, Twentynine Palms, CA across three days. Each FiST team completed two hour-long simulation runs in which they were presented with a subset of three possible mission scenarios presented using the Xiphos JTAC Virtual Trainer (JVT) simulation system (*JTAC Virtual Trainer (JVT)*, 2024). The mission scenarios included three difficulty levels: *Scenario 1*, with the FiST in a defensive position responding to a surprise attack; *Scenario 2*, with the FiST in an offensive position supporting maneuvering forces; and *Scenario 3*, with the FiST in an offensive position supporting a platoon against increased enemy capabilities. Three officer level instructors playing the role of ground force command, air command, and indirect fires were positioned behind the FiST, observing and rating the team's mission success (Mission Rating Score) based on a subjective scoring rubric developed by the NHRC and Naval Air Warfare Training Systems Division (NAWCTSD) (**Table 2**).

Table 2: Overall Mission Scoring

Mission Score	Outcome	Execution	Impact
1 – Mission Failure	The overall commander's intent was not met, and none or very few objectives were achieved.	Several major mistakes, errors, or delays significantly hindered progress or made success impossible.	Mission failed to deliver meaningful results or achieve its core purpose.
2- Partial Mission Failure	The overall commander's intent was not met, but some aspects of the mission were achieved	Significant mistakes, errors, or delays occurred, limiting success and preventing full mission accomplishment.	While aspects of the mission succeeded, the failures overshadowed the partial successes
3 - Mission Success with Notable Issues	The overall commander's intent was met, but execution was inefficient or suboptimal.	Mistakes, delays, or errors occurred that noticeably impacted effectiveness but did not prevent success.	The mission succeeded but left room for improvement in performance.
4 - Mission Success	The overall commander's intent was fully met in an effective and efficient manner.	Few or no impactful mistakes, delays, or errors occurred. Any issues were minor or quickly resolved.	The mission was executed at a high standard, demonstrating strong planning and execution.

Data Acquisition and Procedures

The FiST was set up with five team members sitting parallel each interacting with the simulation software and outfitted with a B-alert X24 EEG headset and high-resolution tEEG (Figure 2). As described above, each of the four FiSTs completed two simulation runs that were roughly 60 minutes long for a total of 8 missions. Recording occurred across a four-day period at Twentynine Palms.

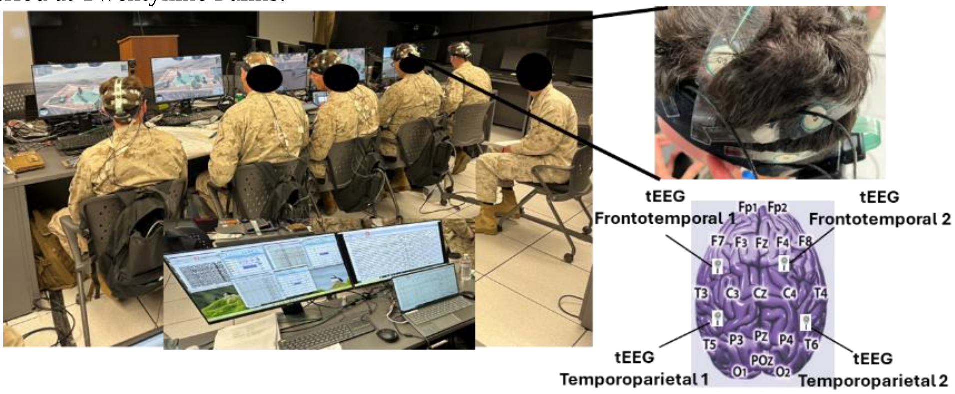

Figure 2: (left) Five member FiST team from left to right JTAC, FiST Lead, FSO, FO Artillery, FO Mortar with instructor seated behind the team. Zoomed in view of recording station is overlaid at bottom of image. (right, top) Zoomed in view of B-alert X24 EEG cap with tEEG electrodes added (right, bottom) B-alert X24 electrodes positioned as subset of 10-20 system with added frontotemporal and temporoparietal positioning of tEEG electrodes

Neural Recordings

Neural EEG recordings were made using mobile, wireless B-Alert® X24 system. This EEG system has been deployed previously in DoD training environments and uses a lightweight, comfortable form factor supporting ease of application and use with sensor preparation and calibration requiring 10 minutes plus a one-time, brief 9-minute benchmark session to match derived calibration models to an individual. The cap consists of 20 electrodes on the scalp (Fp1, Fp2, F7, F8, F3, F4, Fz, T3, T4, C3, C4, Cz, T5, T6, P3, P4, Pz, POz, O1, O2), a subset of 10-20 system with reference electrodes in linked mastoid positions and data digitized at 256 Hz.

Additionally, to more precisely pinpoint network level brain activity related to team interactions from deeper brain regions, we combined the B-alert® with tripolar concentric ring electrodes (tCREs) (Figure 2). These specialized tEEG electrodes provide a higher resolution measure than traditional EEG sensor systems and combat the effects of low spatial resolution caused by spatial aliasing/smearing from the effect of the brain meninges, skull, and scalp distorting source localization (Srinivasan et al., 1998). tCREs are passive electrodes that counter the effects of spatial aliasing by using measurements from multiple sensor rings to provide an approximation of the Laplacian or second order spatial derivative during measurement (Besio et al., 2006). This effectively generates a current source density measurement and increases the spatial resolution, signal-to-noise ratio (SNR) and the ability to identify candidate neural signals from deeper brain regions. Four tEEG electrodes were positioned in frontotemporal and temporoparietal positions, with a reference electrode at the left mastoid and data digitized at 2048 Hz. tEEG electrode positions were chosen to not interfere with B-alert electrode placement while also targeting 10-20 electrode positions.

Peripheral Physiologic Recordings and Audio/Video

A computer monitor mounted Tobii eye tracker nano (sampled at 50 Hz) was used to track eye movements across the JVT simulation monitor. The B-Alert® single channel ECG set up was used for R-R spike identification to allow for both the measure of beat-to-beat heart rate (HR) and heart rate variability (HRV) (Berka et al., 2004, 2007; Johnson et al., 2011). Individual web cameras (30 frames/sec) were positioned on each team member in addition to a camera covering the full FiST team. Audio lapel mics (44.1 kHz) were also worn and connected to an external system allowing simulated radio communications between team members and the instructors playing the roles of ground force command and air command.

All computing resources were synced to an external Network Time Protocol (NTP) server with a NTP Daemon (NTPD) local time service running in the background on each recording and simulation machine to ensure consistent sub-millisecond time synchronization.

Data Cleaning and Analysis

Data cleaning procedures included post-processing for artifact reduction using MNE Python (Larson et al., 2024). This approach was used in conjunction with the B-Alert® online artifact removal which applies early-stage denoising to remove spikes, amplifier saturation, excursions, eye blinks or excessive muscle movement in real time with interpolation for denoised or missing data. Offline artifact removal included:

- Visual inspection of artifacts
- Notch filter at 60Hz to remove line noise
- ICA for Eye-blink removal using Fp1 and Fp2 channels to detect vertical eye movements (VEOG)
- ICA for Muscle artifact removal
- B-Alert® EEG bandpass filtered 1-40Hz
- tEEG bandpass filtered 1-200Hz

Derived functional connectivity metrics for measuring interbrain connectivity between dyads of the FiST including phase-locking value (PLV) and spectral coherence (COH) (Equation 1 and 2 below) were calculated using the HyPyP, a hyperscanning Python pipeline for interbrain connectivity analysis (Ayrolles et al., 2021). Metrics were computed as interbrain measures and averaged across all combinations of dyads (10 total per team) of the five-member FiST to establish a team-level connectivity measure. These two approaches were chosen as well-validated measures of inter-brain coupling allowing for the distillation of a complex web of pairwise relationships into a single, robust team-level metric that could be linked to behavioral or performance outcomes (Dikker et al., 2017; Poulsen et al., 2017). Additional measures such as engagement and workload for individual team members was also provided by the B-Alert® system. Dwell time along with time away from screen was also calculated based on the eye tracker data.

$$\boldsymbol{Phase\ Locking\ Value\ (PLV)_{ij}} = \left|(\mathbf{1}/\boldsymbol{N})\boldsymbol{\Sigma}_{\boldsymbol{n=1}}^{\boldsymbol{N}}\boldsymbol{e}^{\boldsymbol{i}\left(\boldsymbol{\varphi_i(n)}-\boldsymbol{\varphi_j(n)}\right)}\right| \qquad \textbf{(1)}$$

$$\boldsymbol{Spectral\ Coherence\ (COH)_{xy}(f)} = \left|\boldsymbol{S_{xy}(f)}\right|^{\mathbf{2}} / \left(\boldsymbol{S_{xx}(f)S_{yy}(f)}\right) \qquad \textbf{(2)}$$

Causal model development

To work towards building a predictive model of team performance that reflects the theory and the data, we used top-down hypothesis driven testing of model features. The idea is to *let theory guide machine learning* to find the best model by using expert knowledge of TBCT and the neural underpinnings of teaming. Using combinations of behavioral, neural and peripheral physiologic signals we can make inroads into which determining signals are necessary and sufficient to form a cohesive signature predictive of team performance (Figure 3).

In developing a causal model for predicting subjective team scoring, we leverage a set of pre-identified top-down neural features—specifically COH and PLV measured across discrete windows of team behavioral activity aligned to key task events. Our two target behavioral windows included: 1) the *Mission Brief*, which occurs at the start of each 1-hour mission, and involved an instructor providing directions on the details of the scenario configuration; and 2) *Attention to FiST*, which is a specific call to attention by which FiST members alert the other team members that key information is being passed regarding the action for fire support and can occur multiple times throughout the 1-hour mission. Within a structural causal framework, we posit directed influences from moment-to-moment neural synchrony metrics within these key behavioral windows to aggregated subjective performance ratings provided by instructors (see **Table 2**).

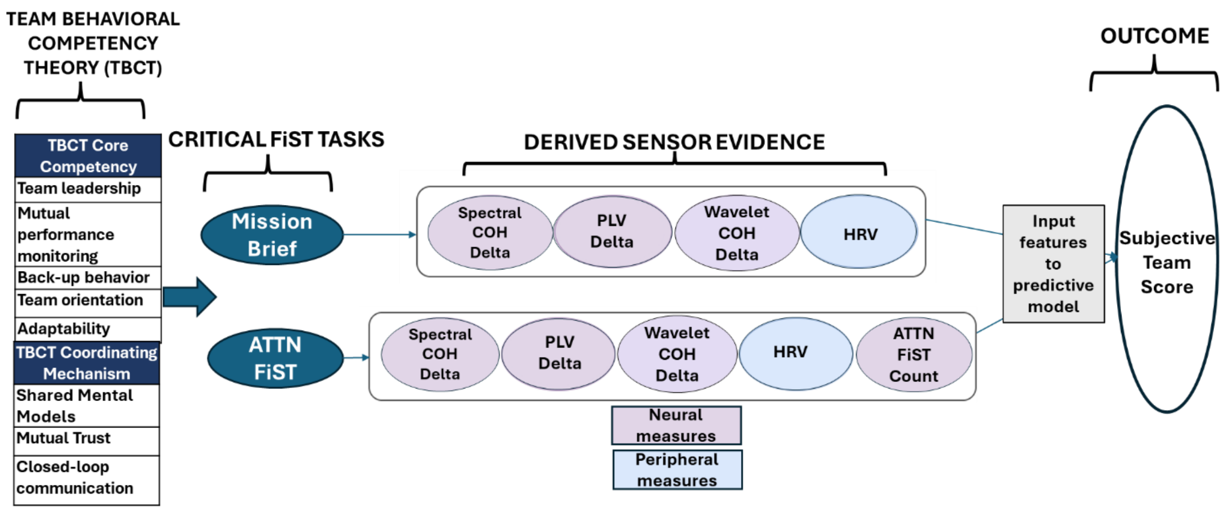

Figure 3: Initial causal model driven by TBCT linked to critical FiST Tasks, derived sensor evidence and team performance outcomes

To control for confounds such as individual baseline arousal or state, we test two hypotheses by comparing our *main windows* of activity (*Mission Brief and Attention to Fist*) to a *pre-stimulus window* set as a period prior to the start of the mission (Figure 4). Hypothesis 1 (H1) tests if a greater delta in PLV across the team between *pre-stimulus* and *Mission Brief* and *Attention FiST* windows correlates with a higher subjective mission rating. Hypothesis 2 (H2) tests if a greater delta in COH connections across the team between pre-stimulus and *Mission Brief* and *Attention FiST* windows correlates with a higher subjective mission rating. For PLV we directly calculate a difference between the two windows (pre-stimulus and main). For COH we establish a threshold by inspection and calculated the difference in connections above threshold for each window.

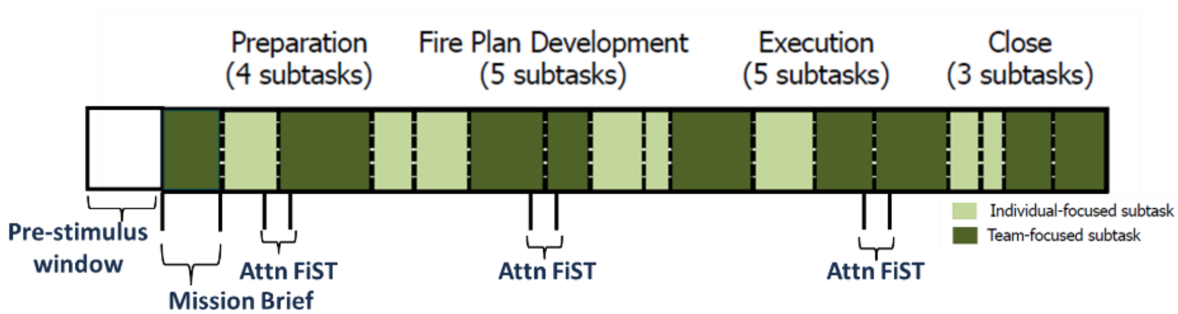

Figure 4: Example windows of activity during the mission, displaying a pre-stimulus window and behavioral windows such as Mission Brief and Attention to FiST as part of a notional FiST task overview.

Feature extraction yielded both windowed averages and transient peaks of coherence and PLV, which are normalized and de-meaned to capture sustained versus phasic inter-brain coupling (Hakim et al., 2023). To test each feature's relevance and populate the causal model, we compute simple pairwise correlations between each synchrony feature and the aggregated team score to identify strong associations. As we develop our causal model, we include only those edges in the graph whose correlation coefficients exceed a predetermined threshold, using these values as initial path weights.

RESULTS

Our initial results focus on testing the ability to develop a causal model grounded in TBCT and specifically examine neural connectivity features that drive teaming processes. In testing Hypothesis 1, our results supported the hypothesis and showed that the change in PLV across the team between *pre-stimulus* and both the *Mission Brief* and *Attention FiST* windows correlated with a higher subjective mission rating, specifically in the delta and beta frequency bands. In testing Hypothesis 2, our results supported the hypothesis and showed that the change in the number of significant COH connections across the team between the pre-stimulus and the *Attention FiST* window correlated with a higher subjective mission rating, specifically in the delta frequency band. These results are displayed in Figure 5 as a series of correlation coefficients that were calculated across each of the target behavioral windows. (*Mission Brief and Attention to FiST*) for the neural functional connectivity metrics spectral coherence and phase-locking value against subjective mission rating scores (n=8, four total FiSTs with two simulation runs per team). We focus on specific known EEG frequency bands: delta (1-3 Hz), theta (3-7 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (30-40 Hz). Figure 5 shows initial results for the correlation between Mission rating and both EEG and tEEG functional connectivity metrics across specific behavioral windows (W), frequency bands (F), and sensor type (S). Note that for EEG we averaged across all 20 electrode positions and for tEEG across all four tEEG electrode positions.

The results show the highest correlation values for the Attention to FiST Window for PLV, and COH calculated across the delta frequency range for EEG (r = 0.70 , p = 0.08) (r = 0.50 , p = 0.25), respectively. For tEEG the highest correlation value was found for Attention to FiST Window for COH (r = 0.45 , p = 0.31).

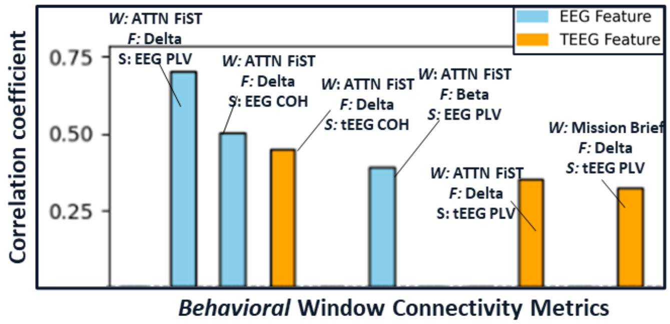

Figure 5:Highest correlation coefficients calculated across Attention to FiST and Mission Brief Windows for PLV and COH for targeted frequency windows in process of building up feature inputs to causal model

To focus in more detail on each of these specific features we show the underlying scatter plots for each feature (Figure 6 and Figure 7). The r values in each scatter plot match the height of the corresponding bar graph correlation results in Figure 5, and p values are reported.

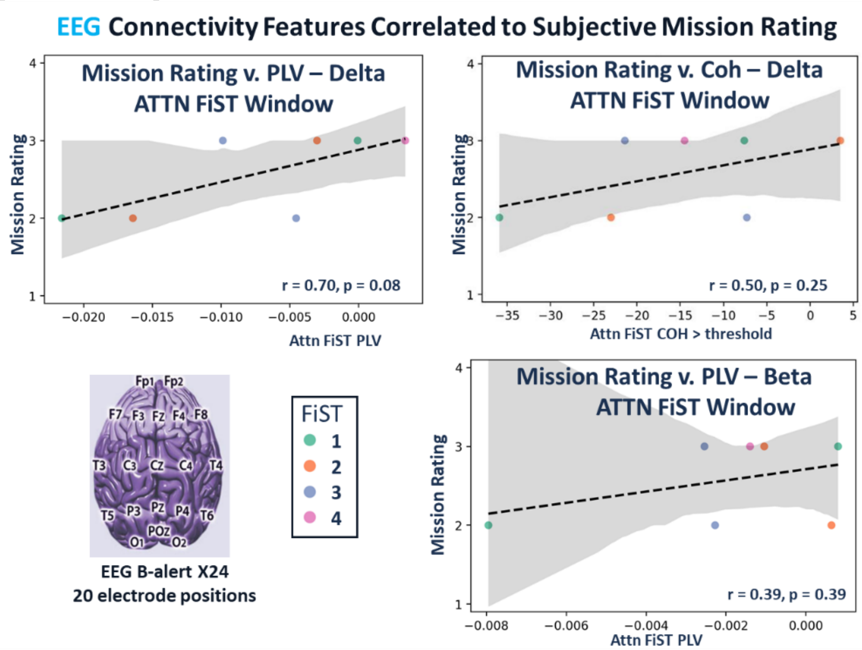

Figure 6: Scatter plots for 4 FiSTs (Teams 1-4), two simulation runs per team, showing the highest correlation value results for EEG connectivity features as displayed in the light blue bars in Figure 5. Plots are all specific features defined by functional connectivity measure (PLV or COH), Frequency range (delta, theta, alpha, beta, gamma). Linear regression fit r and p values are reported in the lower right-hand corner of the plot.

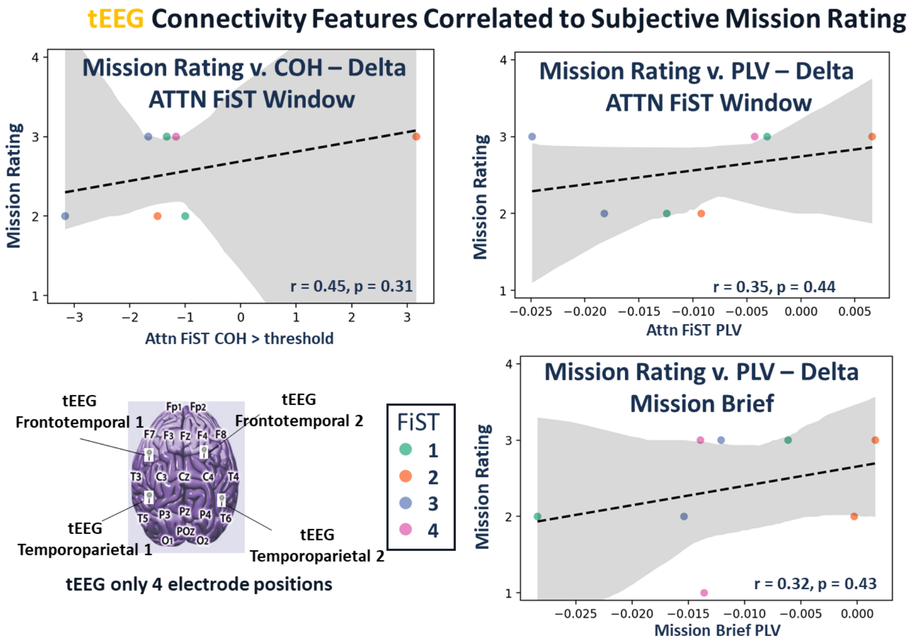

Figure 7: Scatter plots for 4 FiSTs (Teams 1-4), two simulation runs per team, showing the highest correlation value results for tEEG connectivity features as displayed in the light orange bars in Figure 5 . Plots are all specific features defined by functional connectivity measure (PLV or COH), Frequency range (delta, theta, alpha, beta, gamma). Linear regression fit r and p values are reported in the lower right-hand corner of the plot.

DISCUSSION AND CONCLUSIONS

We present our initial results for n = 8 mission simulation runs across four FiSTs. Given the small sample size we plan for further data collects with the goal to expand to n = 24 total mission simulation runs across two more data collection events. With that said, our initial results support our hypotheses that greater team neural coordination - as measured by Phase-Locking Value (PLV) and Spectral Coherence (COH) within specific frequency bands e.g., delta band - correlates with higher Mission Ratings providing guidance into understanding some of the key features needed to build out a causal model for predicting subjective scores of team performance based on objective neural data. Our approach is grounded in Team Behavioral Competency Theory, which allows us to better link functional connectivity metrics between team members with not only critical FiST specific tasks but additional broader team-level interactions that drive team dynamics.

What is noticeable is the dominance of delta-band synchrony during our targeted behavioral windows. Delta-band synchrony often peaks during key coordinated behaviors because slow oscillations naturally integrate information over longer timescales, matching the duration of complex team interactions (Chuang et al., 2024; Lakatos et al., 2008; Réveillé et al., 2024; Szymanski et al., 2017). In our study, both the Attention to FiST and Mission Brief windows demand joint attention, shared decision-making, and potentially rapid error monitoring, driving each member's delta rhythms into phase and amplitude alignment. This low-frequency coupling is both robust to transient noise and sensitive to the salience of collective events, so when teams feel most "in sync" subjectively—navigating a critical task or responding to an unexpected cue—their delta coherence and PLV both rise. In short, delta synchrony provides a neural substrate for group-level integration, and it best captures those slower, high-order processes that underpin teams' shared sense of performance.

Additionally, we note that in testing Hypothesis 1 (greater change in PLV across the team between *pre-stimulus* and *Mission Brief* and *Attention FiST* windows correlates with a higher subjective mission rating) and Hypothesis 2 (a greater change in COH across the team between pre-stimulus and *Mission Brief* and *Attention FiST* windows correlates with a higher subjective mission rating) we observe negative values on the x-axis. This implies that our pre-stimulus baseline window had larger connectivity values than our main behavioral window. This result is contrary to our expectation of increased synchronization during the main window. This suggests that rather than an increase in connectivity there is a maintenance of synchronization seen in higher performing teams and a loss of synchronization that occurs for poor performing teams. The linear relationship between mission rating and the underlying connectivity feature remains. In other words, it is the consistency of synchronization that appears to drive team performance.

Limitations

Despite the above implications there are significant limitations to drawing broader conclusions regarding feature development for causal predictions. To start, our p-values are unreliable given the lower statistical power due to the limited sample size and likelihood for high random sampling variability. Also, in terms of neurodynamics it is quite plausible that elevated delta-band PLV and COH simply reflect a shared, stimulus-locked slow wave (an evoked potential) rather than genuine inter-brain coupling. When the full team is presented with the same cue, their cortices produce a large, stereotyped slow-wave response (e.g. a contingent negative variation or movement-related potential) that is generated in the delta band. Since this evoked activity is time and phase-locked across participants, it inflates zero-lag synchrony metric computations. In other words, the result could be due to the collective reaction to the event or an artifact of simultaneous stimulus processing instead of true neural synchronization. To mitigate this potential implication, we are working to expand our feature identification process as described below.

As we work to build out the causal model using a top-down feature identification approach, we have observed difficulties in covering the feature space. It is possible that we are omitting higher-order combinations, missing hidden drivers of neural sync, overfitting to current data, or mis-quantifying complex constructs. To address these limitations, we are building out an approach that combines top-down subject matter expert guided features with bottom-up data-driven discovery, e.g., automated feature selection, dimensionality reduction, and higher dimensional manifolds. This hybrid approach can balance theoretical rigor with empirical robustness. The idea is that bottom-up approaches can identify team neural states analogous to EEG microstates observed at the individual level (Mishra et al., 2020). This would imply that the oscillatory neural activity of team members enters repeated synchronized team neural states that drive team communication and interactions and result in higher team performance.

Future Work

Future efforts will include increasing our sample size with more data collection events and working to improve the signal-to-noise ratio not only in our raw data but in our derived feature and behavioral window comparisons. This includes using a more suitable pre-stimulus window and testing our current and future recordings with sham or surrogate teams created by calculating functional connectivity metrics across pairs of team members performing simulation runs at different time periods. For example, calculating functional connectivity metrics for team member 1 on mission/run 1 with team member 2 on mission/run 2, separated in time. We also plan to flesh out our causal model with increased top-down hypotheses including our peripheral physiologic recordings such as HRV and eye-tracker gaze and dwell time. This will further complement our additional planned analysis using bottom-up, data-driven discovery methods that apply nonlinear dynamic approaches to analyze team behavioral and neural trajectories to determine how dynamical properties of team cluster time courses (team neural states) distinguish between high and low team performance.

By using TBCT to ground our approach in team theory, we set the stage for practical scalability and generalizability across other DoD teams and settings. The output will be a transparent, explainable model, open to inspection to improve trust and understanding in objective team performance predictions. We are in the nascent stages of building our causal model, but theoretical grounding provides the basis for which to map domain-specific tasks to more generalized team performance to build out an extra toolset for instructors to supplement current subjective measures of team success. The target result is a scalable and interpretable framework for assessing DoD team readiness.

ACKNOWLEDGEMENTS

This work was completed under DARPA contract #HR001124C0497 as part of the Objective Prediction of Team Effectiveness via Models of Performance Outcomes (OP TEMPO) program. The authors would like to acknowledge the DARPA team, the government testing and evaluation team from NHRC, NAWCTSD and Air Force Research Labs (AFRL), and the Twentynine Palms Marine Ground Air Combat Center (MGACC) Battle Simulation Center leadership and FiST participants whose expertise, day-to-day assistance and participation were invaluable to this work.

REFERENCES

Acquadro, M. A. S., Congedo, M., & De Riddeer, D. (2016). Music Performance As an Experimental Approach to Hyperscanning Studies. *Frontiers in Human Neuroscience*, *10*, 242. https://doi.org/10.3389/fnhum.2016.00242

Antonenko, P. D., Davis, R., Wang, J., & Celepkolu, M. (2019). On the Same Wavelength: Exploring Team Neurosynchrony in Undergraduate Dyads Solving a Cyberlearning Problem With Collaborative Scripts. *Mind, Brain, and Education*, *13*(1), 4–13. https://doi.org/10.1111/mbe.12187

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., He, E. J., Roy, A., Cincotti, F., Salinari, S., Mattia, D., He, B., & Babiloni, F. (2012). Cortical activity and functional hyperconnectivity by simultaneous EEG recordings from interacting couples of professional pilots. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2012*, 4752–4755. https://doi.org/10.1109/EMBC.2012.6347029

Ayrolles, A., Brun, F., Chen, P., Djalovski, A., Beauxis, Y., Delorme, R., Bourgeron, T., Dikker, S., & Dumas, G. (2021). HyPyP: A Hyperscanning Python Pipeline for inter-brain connectivity analysis. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 72–83. https://doi.org/10.1093/scan/nsaa141

Barraza, P., Pérez, A., & Rodríguez, E. (2020). Brain-to-Brain Coupling in the Gamma-Band as a Marker of Shared Intentionality. *Frontiers in Human Neuroscience*, *14*. https://www.frontiersin.org/articles/10.3389/fnhum.2020.00295

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. (2004). Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. *International Journal of Human–Computer Interaction*, *17*(2), 151–170. https://doi.org/10.1207/s15327590ijhc1702_3

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B231-244.

Berka, C., & Stikic, M. (2017). On the Road to Autonomy: Evaluating and Optimizing Hybrid Team Dynamics. In W. F. Lawless, R. Mittu, D. Sofge, & S. Russell (Eds.), *Autonomy and Artificial Intelligence: A Threat or Savior?* (pp. 245–262). Springer International Publishing. https://doi.org/10.1007/978-3-319-59719-5_11

Besio, W. G., Koka, K., Aakula, R., & Dai, W. (2006). Tri-polar concentric ring electrode development for laplacian electroencephalography. *IEEE Transactions on Bio-Medical Engineering*, *53*(5), 926–933. https://doi.org/10.1109/TBME.2005.863887

Chuang, C.-H., Peng, P.-H., & Chen, Y.-C. (2024). *Leveraging Hyperscanning EEG and VR Omnidirectional Treadmill to Explore Inter-Brain Synchrony in Collaborative Spatial Navigation* (arXiv:2406.06327). arXiv. https://doi.org/10.48550/arXiv.2406.06327

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology*, *27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Hakim, U., De Felice, S., Pinti, P., Zhang, X., Noah, J., Ono, Y., Burgess, P. W., Hamilton, A., Hirsch, J., & Tachtsidis, I. (2023). Quantification of inter-brain coupling: A review of current methods used in haemodynamic and electrophysiological hyperscanning studies. *NeuroImage*, *280*, 120354. https://doi.org/10.1016/j.neuroimage.2023.120354

Hoehl, S., Fairhurst, M., & Schirmer, A. (2021). Interactional synchrony: Signals, mechanisms and benefits. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 5–18. https://doi.org/10.1093/scan/nsaa024

Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., & Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biological Psychology*, *133*, 54–62. https://doi.org/10.1016/j.biopsycho.2017.12.005

Jahng, J., Kralik, J. D., Hwang, D.-U., & Jeong, J. (2017). Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game. *NeuroImage*, *157*, 263–274. https://doi.org/10.1016/j.neuroimage.2017.06.024

Johnson, R. R., Popovic, D. P., Olmstead, R. E., Stikic, M., Levendowski, D. J., & Berka, C. (2011). Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. *Biological Psychology*, *87*(2), 241–250. https://doi.org/10.1016/j.biopsycho.2011.03.003

JTAC Virtual Trainer (JVT): Scalable VR Training Solution. (2024, August 29). https://xiphos.partners/jvt/

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.)*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Larson, E., Gramfort, A., Engemann, D. A., Leppakangas, J., Brodbeck, C., Jas, M., Brooks, T. L., Sassenhagen, J., McCloy, D., Luessi, M., King, J.-R., Höchenberger, R., Goj, R., Brunner, C., Favelier, G., van Vliet, M., Wronkiewicz, M., Rockhill, A., Holdgraf, C., … luzpaz. (2024). *MNE-Python* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.14519545

Mishra, A., Englitz, B., & Cohen, M. X. (2020). EEG microstates as a continuous phenomenon. *NeuroImage*, *208*, 116454. https://doi.org/10.1016/j.neuroimage.2019.116454

Pan, Y., Dikker, S., Goldstein, P., Zhu, Y., Yang, C., & Hu, Y. (2020). Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage*, *211*, 116657. https://doi.org/10.1016/j.neuroimage.2020.116657

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific Reports*, *7*(1), 43916. https://doi.org/10.1038/srep43916

Reinero, D. A., Dikker, S., & Van Bavel, J. J. (2021). Inter-brain synchrony in teams predicts collective performance. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 43–57. https://doi.org/10.1093/scan/nsaa135

Réveillé, C., Vergotte, G., Perrey, S., & Bosselut, G. (2024). Using interbrain synchrony to study teamwork: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *159*. https://doi.org/10.1016/j.neubiorev.2024.105593

Salas, E., Sims, D. E., & Burke, C. S. (2005). *Is there a "Big Five" in Teamwork?* https://journals.sagepub.com/doi/10.1177/1046496405277134.

Sciaraffa, N., Borghini, G., Aricò, P., Di Flumeri, G., Colosimo, A., Bezerianos, A., Thakor, N. V., & Babiloni, F. (2017). Brain Interaction during Cooperation: Evaluating Local Properties of Multiple-Brain Network. *Brain Sciences*, *7*(7), 90. https://doi.org/10.3390/brainsci7070090

Srinivasan, R., Tucker, D. M., & Murias, M. (1998). Estimating the spatial Nyquist of the human EEG. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 8–19. https://doi.org/10.3758/BF03209412

Stevens, R., Galloway, T., Berka, C., & Behneman, A. (2010). *A Neurophysiologic Approach For Studying Team Cognition*.

Stevens, R., Galloway, T., Wang, P., & Berka, C. (2011). Mapping Neurophysiologic Synchrony Attractor States and Entropy Fluctuations during Submarine Piloting and Navigation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 212–216. https://doi.org/10.1177/1071181311551044

Stevens, R., Galloway, T., Wang, P., Berka, C., Tan, V., Wohlgemuth, T., Lamb, J., & Buckles, R. (2013). Modeling the neurodynamic complexity of submarine navigation teams. *Computational and Mathematical Organization Theory*, *19*(3), 346–369. https://doi.org/10.1007/s10588-012-9135-9

Stevens, R. H., & Galloway, T. L. (2016). Modeling the neurodynamic organizations and interactions of teams. *Social Neuroscience*, *11*(2), 123–139. https://doi.org/10.1080/17470919.2015.1056883

Stevens, R. H., Galloway, T. L., Wang, P., & Berka, C. (2012). Cognitive neurophysiologic synchronies: What can they contribute to the study of teamwork? *Human Factors*, *54*, 489–502. https://doi.org/10.1177/0018720811427296

Stikic, M., Berka, C., Waldman, D., Balthazard, P., Pless, N., & Maak, T. (2013). Neurophysiological Estimation of Team Psychological Metrics. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition* (pp. 209–218). Springer. https://doi.org/10.1007/978-3-642-39454-6_22

Stone, B., Skinner, A., Stikic, M., & Johnson, R. (2014). *Assessing Neural Synchrony in Tutoring Dyads*. 167–178. https://doi.org/10.1007/978-3-319-07527-3_16

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The Five Myths of MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, *27*(4), 553–564. https://doi.org/10.1007/s10548-013-0326-6

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., Müller, V., & Lindenberger, U. (2017). Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *NeuroImage*, *152*, 425–436. https://doi.org/10.1016/j.neuroimage.2017.03.013

Zajac, S., Holladay, C. L., Tannenbaum, S., & Salas, E. (2022). Building effective healthcare team development interventions in uncertain times: Tips for success. *Organizational Dynamics*, *51*(2), 100824. https://doi.org/10.1016/j.orgdyn.2020.100824

Zhang, M., Jia, H., & Wang, G. (2021). Interbrain Synchrony of Team Collaborative Decision-Making: An fNIRS Hyperscanning Study. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.702959