Proceedings of the Human Factors and Ergonomics Society Annual Meeting
2025, Vol. 69(1) 1255–1259

DOI: 10.1177/10711813251369877
journals.sagepub.com/home/pro
Sage

BioTDMS: A Multi-Sensor and Data-driven System for Real-Time Team Performance Assessment

Elmira Zahmat Doost[1], Jamie C. Gorman[1], Xiaoyun Yin[1], Shiwen Zhou[1], Garima Arya Yadav[1], Nancy Cooke[1], Lixiao Huang[1], Bethany K. Bracken[2], Aaron Winder[2], Phillip Desrochers[2], Mengyao Li[3], Emanuel Rojas[3], Michael Rosen[4], and Molly Kilcullen[5]

Abstract

The Biobehavioral Team Dynamics Measurement System (BioTDMS) presents a novel, multimodal approach for assessing real-time team performance in high-stakes, dynamic environments. Addressing the limitations of post hoc, subjective evaluations, BioTDMS integrates physiological (EEG, fNIRS, ECG, respiration), behavioral (gaze, speech, input logs), and communication data to generate dynamic, objective metrics of team cognition and adaptability. Grounded in interactive team cognition theory and layered dynamics modeling, the system captures reorganization and synchrony in team states in response to task perturbations. Using data collected from eight dyadic teams (16 participants) performing a time-constrained Noncombatant Evacuation Operation (NEO) planning task, BioTDMS identifies bio-behavioral signatures predictive of team effectiveness and stress resilience. Machine learning models, particularly logistic regression and support vector machines demonstrated high predictive performance (F1 $\approx$ 0.96) when leveraging team-level synchrony features, outperforming models trained on individual-level data alone. These findings underscore the importance of interactional metrics in team performance assessment and offer a pathway toward adaptive, data-driven training systems for mission-critical operations.

Keywords

team performance, machine learning, interactive team cognition, layered dynamics modeling, physiological synchrony

Introduction

Effective teamwork is vital in high-stakes settings like military operations and emergency response. However, current assessments often rely on post hoc, subjective evaluations that fail to capture the dynamic, time-varying nature of team cognition. The Biobehavioral Team Dynamics Measurement System (BioTDMS) addresses this gap by integrating multimodal physiological and behavioral data with explainable machine learning models. Grounded in real-time team cognition (Gorman et al., 2020), BioTDMS captures emergent and reorganizing team states, producing real-time, objective, and generalizable metrics for team adaptability and effectiveness (Figure 1). The system aims to identify, characterize, and validate bio-behavioral signatures of team enaction, adaptation, and recovery (i.e., readiness and resilience), which are essential for optimizing team effectiveness and operational decision-making in dynamic environments. The primary objectives of BioTDMS are: (1) Develop a multimodal sensing system that fuses neural (EEG, fNIRS), physiological (ECG, respiration, PPG), and behavioral (speech, gaze, input logs) signals; (2) Apply layered dynamics modeling to derive adaptive bio-behavioral signatures across individuals and teams; and (3) Validate the predictive performance of machine learning classifiers using synchronized sensor data and ground-truth task scores.

Background

Team dynamics present key challenges in linking individual bio-behavioral signals to team-level processes and outcomes.

[1]Arizona State University, Polytechnic, Mesa, USA
[2]Charles River Analytics Inc, Cambridge, USA
[3]Georgia Institute of Technology, Atlanta, USA
[4]Johns Hopkins Medicine, Baltimore, MD, USA
[5]Johns Hopkins University, Baltimore, MD, USA

Corresponding Author:
Elmira Zahmat Doost, Arizona State University, 1151 S Forest Avenue, Tempe, AZ 85287-1004, USA.
Email: ezahmatd@asu.edu

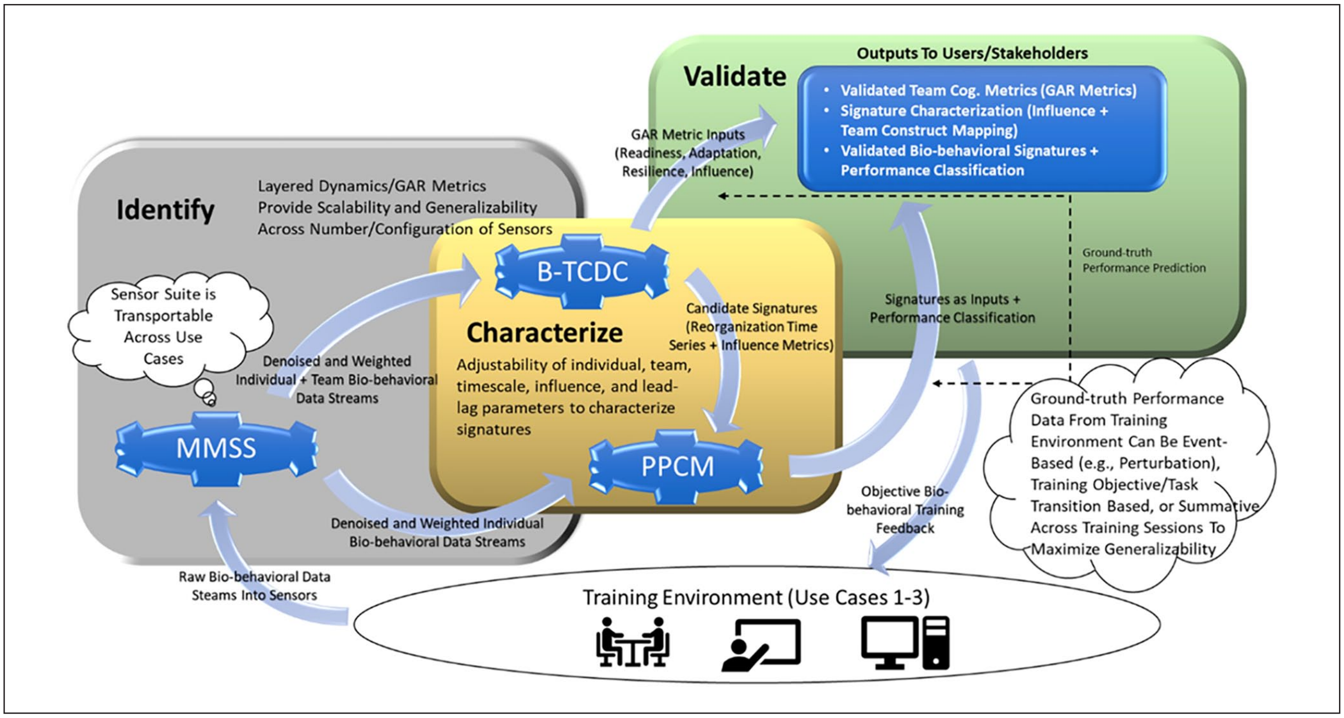

Figure 1. BioTDMS overview.

Traditional models rely on aggregate, static snapshots of team cognition and/or subjective knowledge assessments, which fail to capture the nonstationary, time-varying nature of team interactions in dynamic environments (Gorman & Wiltshire, 2024). Also, many dynamic approaches to team assessment borrow theory from physical dynamical systems (e.g., synchronization, recurrence) that do not fully reflect the complexity of human interaction. Reliance on static physical models is problematic because team coordination is continuously reorganizing across multiple timescales and levels of analysis, necessitating a multiscale (individual and team), multimodal (bio and behavioral), moving window approach to capture dynamic patterns. BioTDMS builds on recent advances in interactive team cognition theory (Cooke et al., 2013) to enable real-time bio-behavioral monitoring, coupled with machine learning-based performance prediction, to provide a generalizable, predictive framework for assessing team effectiveness.

BioTDMS integrates physiological and behavioral data to dynamically map team coordination and adaptability onto "ground-truth" performance metrics commonly used for military training. This approach extends team cognition theory by incorporating traditional approaches (RF, SVM, MLP, KNN) and Multi-Task Multi-Kernel Learning (MTMKL) techniques developed for generalizable team cognition and performance assessment (Li et al., 2022). BioTDMS utilizes layered dynamics modeling (Gorman et al., 2019), allowing for real-time detection of bio-behavioral signatures that predict team adaptability, decision-making efficiency, and stress responses in teams (Gorman et al., 2016). BioTDMS seeks to enhance team training effectiveness and efficiency by complementing subjective observations of trainers with objective metrics and allowing for a greater pace of training to support increased demands for warfighter readiness.

Approach

We recruited eight dyads ($N = 16$ participants), all with graduate-level experience in human factors and cognitive science, to perform a collaborative Noncombatant Evacuation Operation (NEO; Dunbar & Gorman, 2020; Warner et al., 2003) planning task. Each dyad included an Environmental and a Weapons expert coordinating under time constraints. Data were collected over a 35-min session. To examine team reorganization, two perturbations were introduced (Figure 2): (1) a distractor Lego task and (2) a shortened mission deadline. Participants were equipped with synchronized EEG, fNIRS, ECG, respiration, and eye-tracking sensors (Figure 3). Audio and keyboard/mouse activity were also logged.

The process begins with data fusion and signal processing across multiple channels, including EEG, fNIRS, ECG, respiration, eye tracking, and communication. These signals are synchronized using a global reference timeline and preprocessed using filtering and normalization techniques to minimize noise. Physiological data were downsampled to 1 Hz to expedite preprocessing and validate the feature set.

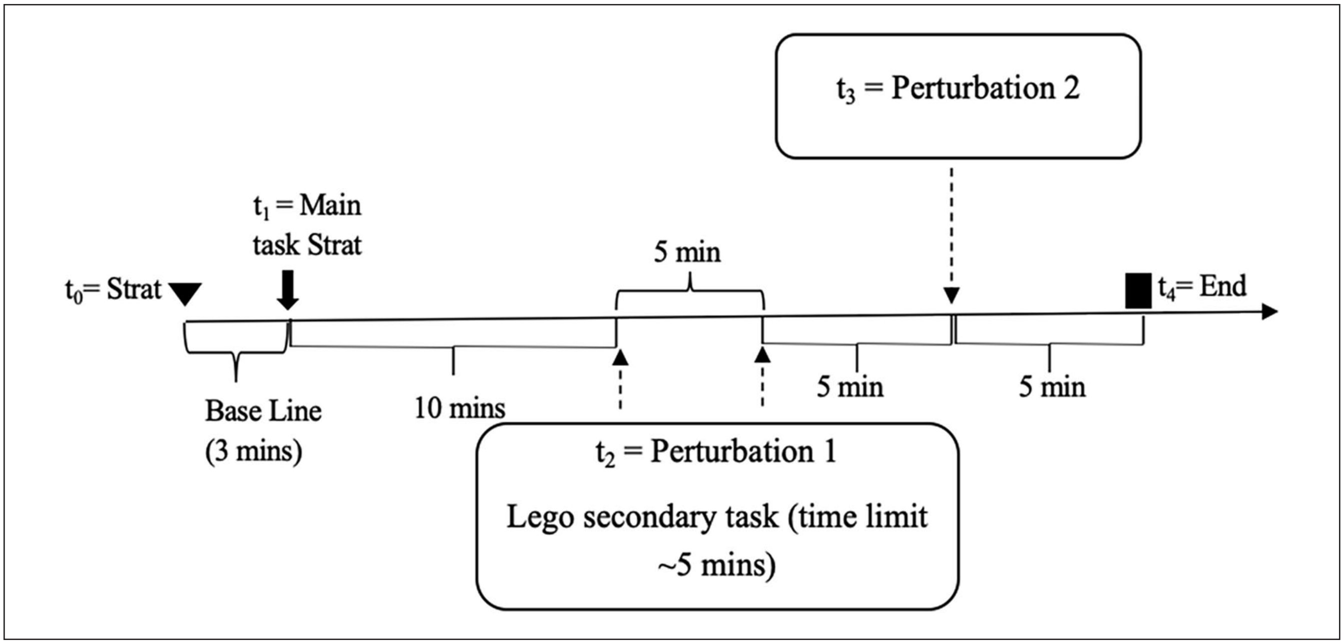

Figure 2. Experimental process.

Figure 3. Experiment set up.

We used a layered dynamics model (Gorman et al., 2019) to generate dynamic reorganization curves of a team across physiological and speech communication sensor. A reorganization time series is computed as the moving window entropy (Shannon & Weaver, 1949) of the team state time series (*ws* = 8 s, #bins = quartiles) (Equation 1). In Equation 1, p_n is the relative frequency of each team communication, vehicle, or physiological state n ∈ N within a time window. The entropy metric was continuously recalculated using a moving-window approach. The informational value (, measured in bits, of an event is weighted by its probability of occurrence. Adaptation is defined as the peak of the reorganization curve extracted from the entropy time series during each perturbation (Stevens & Galloway, 2017). We calculated the 95% confidence interval of the entropy to determine the peak component (Grimm et al., 2023).

$$Entropy = -\sum_{n=1}^{sym} (p_n \times log_2 p_n) \quad (1)$$

Then, to measure the amount of influence (degree to which subsets of sensors account for changing patterns at the team level) we calculated the average mutual information (AMI) between "lower level" (e.g., individual team member; bio-behavioral subsystem) and "higher level" (e.g., team; overall bio-behavioral states) states over time using a moving window of size *ws* (fs = 100 Hz, AMI bins = 10, Lead Lag = 0).

$$AMI_{XY} = \sum_{x_i y_j} P_{XY}(x_i y_j) log_2 \left[\frac{P_{XY}(x_i y_j)}{P_X(x_i) P_Y(y_j)} \right] = Infl(X, Y) \quad (2)$$

Reorganization is used to detect significant shifts in the organization of states across sensors over time, and influence measures the relative importance of different subsystems underlying reorganization. These data provide candidate signature packages that the ML agent uses to optimize performance prediction across both individual and interaction-level.

Initially, 43 features were extracted from multimodal physiological data, including individual-level and interactional-level metrics. To ensure data quality and reduce

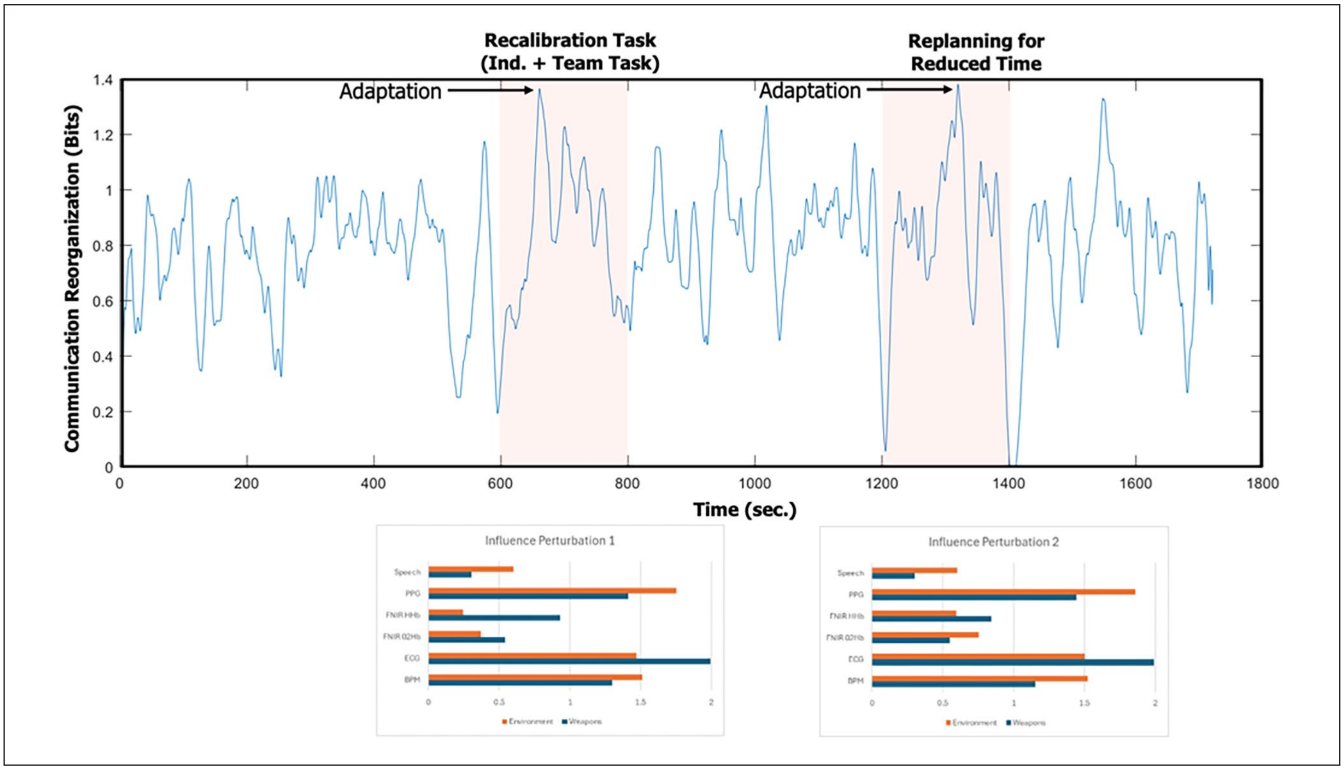

Figure 4. Communication reorganization.

redundancy, we applied filtering and normalization, followed by feature reduction based on three criteria: zero variance, excessive missing values, and high correlation ($r > .7$). This process narrowed the feature set to 22 key features. Individual-level features included heart rate variability (HRV) metrics, respiratory variability, and fNIRS-derived oxygenation trends, while interaction-level features captured team dynamics such as breathing synchrony, photoplethysmogram (PPG) coherence, and oxygenated hemoglobin (O_2HB) synchrony. These features were computed using a sliding window approach—5-min windows for HRV to ensure statistical reliability, and 5-s windows for more dynamic signals. The refined feature set was then used to train and evaluate various models, including traditional approaches (RF, SVM, MLP, KNN) and methods like multi-kernel and multi-task learning. Due to the small sample size and temporal nature of physiological data, we employed a stratified temporal 80 to 20 train-test split.

Outcome

The communication reorganization calculated using the layered dynamics modeling framework across team members is shown in Figure 4. We successfully replicated adaptation peaks in the communication layer (highlighted in pink) across multiple scenarios. The distribution of influence (AMI) was calculated, and the results indicate that the physiological influence distribution is sensitive to perturbations (Figure 5). This finding is important, because changing distributions of bio-behavioral influence can be used to characterize validated signatures sent to the ML models for team performance prediction. We developed our ML pipeline and extracted features using a sliding window approach with two different time scales. Additionally, we computed both individual and interaction-level features, focusing on correlation measures between subjects across physiological data. To improve computational efficiency and model performance, we conducted feature selection. The feature set was then used to train multiple classification models, including KNN, MLP, LR, and SVM. We selected LR for interpretability, SVM for handling non-linear patterns in high-dimensional data, MLP for modeling complex relationships, and KNN as a simple non-parametric baseline. When using only individual-level features, LR and SVM both achieved high performance (F1 $\approx$ 0.86–0.87). Importantly, incorporating interaction-level features—such as O_2Hb coherence and respiratory coupling—boosted the performance of LR and SVM to 96% and 95%, respectively. In contrast, the MLP showed only moderate improvement, while KNN's performance significantly declined, likely due to its sensitivity to high-dimensional feature sets and lack of internal regularization. This range of models enabled us to validate the generalizability of our findings across algorithmic approaches and confirm that team-level synchrony features consistently outperformed individual-only metrics, particularly in interpretable and robust classifiers like LR and SVM.

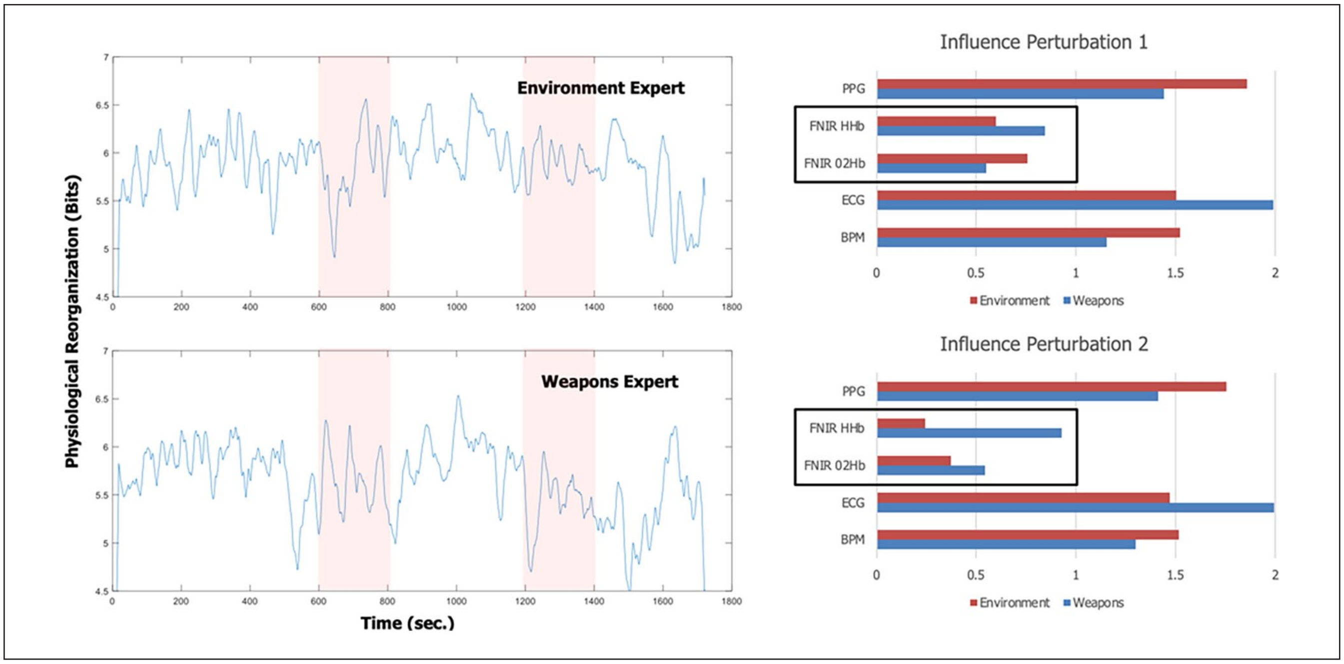

Figure 5. Physiological reorganization.

Conclusion

These findings validate that team-level features are critical for predicting emergent team performance. BioTDMS offers a robust framework for real-time, explainable team performance assessment. Key innovations include multi-resolution metrics that reveal dynamic team reorganization; Influence metrics that identify which subsystems (individual vs. collective) drive performance changes; and High-performing classifiers that highlight the predictive power of team interactional features. Future work includes expanding sample size, validating models across diverse operational domains, and integrating findings into adaptive training interventions aimed at enhancing operational readiness.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is funded by DARPA OP TEMPO FP00041063.

ORCID iD

Elmira Zahmat Doost https://orcid.org/0000-0002-6948-0044

References

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285.

Dunbar, T. A., & Gorman, J. C. (2020). Using communication to modulate neural synchronization in teams. *Frontiers in Human Neuroscience*, *14*, 332.

Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach. *Ergonomics*, *62*(5), 629–643.

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human Factors*, *62*(5), 825–860.

Gorman, J. C., Martin, M. J., Dunbar, T. A., Stevens, R. H., Galloway, T. L., Amazeen, P. G., & Likens, A. D. (2016). Cross-level effects between neurophysiology and communication during team training. *Human Factors*, *58*(1), 181–199.

Gorman, J. C., & Wiltshire, T. J. (2024). A typology for the application of team coordination dynamics across increasing levels of dynamic complexity. *Human Factors*, *66*(1), 5–16.

Grimm, D. A. P., Gorman, J. C., Cooke, N. J., Demir, M., & McNeese, N. J. (2023). Dynamical measurement of team resilience. *Journal of Cognitive Engineering and Decision Making*, *17*(4), 351–382.

Li, M., Erickson, I. M., Cross, E. V., & Lee, J. D. (2022). Estimating trust in conversational agent with lexical and acoustic features. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 544–548.

Shannon, C. E., & Weaver, W. (1949). *A mathematical model of communication* (Vol. 11, pp. 11–20). University of Illinois Press.

Stevens, R. H., & Galloway, T. L. (2017). Are neurodynamic organizations a fundamental property of teamwork? *Frontiers in Psychology*, *8*, 644.

Warner, N., Wroblewski, E., & Shuck, K. (2003). *Noncombatant evacuation operation scenario*. Naval Air Systems Command, Human Systems Department (4.6).