

Fundamental Requirements for Contemporary Performance Evaluation of Emergency-Response and Safety-Critical Teams

Rakesh Veerabhadrapa¹, Imali T. Hettiarachchi¹, Asim Bhatti¹

Abstract—Teams of emergency response and safety-critical (ER&SC) environments are constantly confronted with cognitively challenging tasks demanding a strong interdependence from team individuals. The interdependences are reflected in team attributes such as cooperation, collaboration, joint activity, and aggregates of team cognition. Exploring the impact of attributes on performance can have several benefits in advancing team training methods, identification of best practices of teams and forecasting any critical outcomes. For this purpose physiological and behavioural (PB) techniques have been researched to unravel the relationships between attributes of teams and performance outcomes. PB-based team studies employing single modality have made significant progress over the last two decades. In recent years, this trend has been expanding via the integration of multiple modalities that could revolutionise team studies. Hence, the proposed article selectively reviews team studies employing unimodal and multimodal PB techniques. In the light of input-processes-output team study framework, this article identifies current trends in PB technologies, qualitative and quantitative methods for team performance evaluation, team task environment and performance markers of teams. Further, the article briefly discusses existing methods and technologies with examples of how they were employed in work team settings, as well as, the state of multimodality in the context of teams. Finally, the article highlights some of the gaps in this research domain and provides recommendation on potential future prospects that could better address the gaps. Teams of ER&SC domains can benefit from this review to develop futuristic training and evaluation methods and can potentially enrich teams and their performance.

Index Terms—team performance, physiology & behaviour, interaction dynamics, sensors and technologies, multimodal trends, data processing, data fusion, applied frameworks

I. INTRODUCTION

In this technologically revolutionising era, challenges in work environment are complicated, exceeding the cognitive capacity of individuals [1]. Often, successful achievement of such challenges can be critical, demanding interdependence with more than one party [2, 3], referred to as *teamwork*. Complex challenges are pervasive in emergency response and safety-critical (ER&SC) environments such as clinical surgery, emergency paramedics, military, paramilitary, aviation, construction and nuclear power stations [2, 4, 5, 6]. Capabilities (e.g., efficiency and effectiveness) of teams to accomplish the desired task objectives critically impacts the health, safety, and well-being of many lives [7, 8]. With the future embarked on constant

technological revolution, the reliance on effective performance of teams is bound to exceed more than ever.

The performance of teams is primarily dependent on individuals behaviours (e.g., skills that dictate the capacity of an individual to achieve desired task objectives) and team members' interaction dynamics (e.g., the degree of effective collaboration, coordination, and collective knowledge) [9, 8, 10, 11]. Carefully examining the interactions among team individuals could provide insight into the influential attributes such as affect, behavioural and cognitive states [12]. Such an approach can help understanding intra-team relationships as well as factors influencing at individual level that lead to different levels of performance outcome [10, 13].

Over the last two decades, the science of teams has gained significant momentum [14, 15, 16] towards understanding the factors that influence the performance of teams [10]. Evaluating team performance is paramount for assessing team training outcomes, enhancing team training methods and identifying effective practices of teams, eventually contributing to advancing team functioning [17, 10]. Such advancements can significantly benefit operations of ER&SC teams.

Conventionally, team studies analyse subjective ratings to unfold the impact of influential factors in relation to performance outcome [18, 19, 20]. Subjective ratings are collected from team members (self-reported) and aggregated over the team [9, 21]. However, individuals memory (long or short term) and intent of participation can expose subjective ratings to potential biases [22, 9, 23]. As an alternative, physiological and behavioural recording (PBR) techniques have been extensively researched to unfold complex dynamics of team individuals, interactions and the task characteristics [24, 25]. PBR techniques offer advantage such as continuous monitoring, computational flexibility, long-term record keeping and minimal human interventions [26, 27]. The continuous data provides better temporal resolution that also enabling enhanced statistical analyses. Most recently, the field of computational intelligence (CI), including paradigms such as artificial neural networks, unsupervised classifiers and forecasting algorithms have been rapidly evolving. Co-jointly with CI, PBR-based team studies could contribute to modelling predictive algorithms that can forecast future performance outcomes of team operation [8, 28, 29, 30, 31].

Eleven different models or frameworks were discussed by Salas *et al.* that have contributed to team studies [4]. The *inputs-processes-outputs* (IPO) framework has been widely accepted among the science of teams [8], especially, in studies

that employ the PBR approach [9]. Typical IPO framework-based applied research experimental setup can be composed of at least one PBR technique, a task environment, post-task questionnaires, analytical methods for data pre-processing and statistical methods [8, 32, 33, 34]. Such an experimental setup is necessary to test and evaluate team related theoretical hypotheses [35].

In recent years, team research has witnessed a steady rise in the number of studies experimenting with combination of PBRs (also referred to as multimodal techniques) [36, 37, 38]. Observing this trend, this review presents insights from team studies employing multimodal PBR approach. This will include a thorough analysis of a multimodal approach and its application with respect to the investigated team-related attributes, theories and hypotheses tested in different studies. In the light of ER&SC PBR-based team research, towards advancing effective evaluation and analysis of team performance, this review briefly addresses: (a) the key constructs of IPO framework, (b) projections of team performances observed in the literature, (c) analyses and procedures that were employed in processing the data and team-level measurements, (d) how PBRs can foster a better deciphering of the relationship between team performances, constructs of IPO framework, statistical aggregations and subjective ratings, and (e) finally, how future team research experimental setup can be shaped. The insights presented in this review can promote newer ideas, strengthen a better understanding of underlying team attributes, and foster robust multimodal approaches for effective evaluation of team performance.

II. IPO FRAMEWORK

The IPO framework can help identification of factors contributing to the overall performance of a team [9, 4, 8, 22, 10]. Understanding IPO framework helps to unravel any interrelationship between the constructs *i.e.* the *inputs*, *processes* and *outputs* that concern team performance analysis. Attributes of teams constituted with in each construct is further elaborated as follows.

The *inputs* to the teams act as stimulus which generally set the condition for activating the relevant processes. Team structure and task structure are primary contributors of the *inputs* block [9]. The team structure concerns anatomical information such as number of members in a team, gender composition of a team [39, 40] and team members readiness (knowledge, skills and attitudes) that defines their experience and exposure [35]. The task structure concerns the task activities such as complexity-levels, phases, and objectives of task [9].

In the context of teams, often there exists an interconnectivity between team member responses referred to as *processes*. The *processes* (also referred to as mediators in some literature) are the fundamental factors that influence the performance of a team. Marks *et al.*, conceptualises team processes based on the period of time while a team is involved in a task as well as between episodes of the task [12]. Accordingly, three different elements contribute to processes which are listed as: (a) *Transition process*, the period of time occurring between trials which is responsible for teams to reflect, analyse and

organize various aspects pertaining to enhancing goal achievement [41], (b) *Action process* constitutes the activities that occur while a trial is in progress which is responsible for coordination, resource monitoring, interaction, communication, and external supporting [41] and (c) *Interpersonal*, appears at all times from the moment a team is formed and deployed. It mainly focuses on managing interpersonal relationships, responsible for maintaining harmony, boosting confidence, motivation and emotional balance[41].

The *outcomes* of a team can be observed in their ability to accomplish task objectives such as effectiveness and efficiency [8]. The outcomes also represent learning in teams *i.e.*, transformation in team members' readiness due to experience gained following completion of task routine. Furthermore, effectiveness variables such as satisfaction, viability, and load experienced by team individuals (e.g., workload, cognitive load and temporal load) [8, 9] influence performance.

In addition, as a result of different processes mentioned above, extensive involvement of team members is reflected in the evolution of newer *team states* [12, 42] such as team potency [43], team situational awareness [2], team stress [44], shared knowledge [3] and the trust, respect and cohesion between team members. The team states are responsible for defining qualities of a team such as member attitudes, motivations, values, cooperation, cognition and collaboration that critically define levels team functioning, and hence, the performance.

III. RAMIFICATIONS FOR TEAM PERFORMANCE EVALUATION

The achievement of task objectives using objective scoring could be used as reference for team performance evaluation [35, 18]. In this regard, the following ramifications of team performance evaluations were observed in the literature.

- *Effectiveness and Efficiency* of teams to achieve task objectives such as task completion time [45, 46, 47], account of errors accumulated by teams during a task [45, 47], differences in the quality of final product [48], teams ability to manage resources during a task [24], and efficiency in tracking [24] are some of the variables that can differentiate teams performances.
- *Team performance developmental activities and implications* such as exploring the relationship between variations in objective scores across trials and the evolution of strategies [49, 50]. The amount of practice required for transforming the level of performance to a different level (enhancement or decline in performance) [51].
- *Expert vs novice teams* describe the composition of team based on performance levels. Similar terminologies could include experienced vs inexperienced teams [52, 51]. For example, it is of an interest to understand the physiological and behavioural differences between experts handling a task complexity relative to novices. Time taken to complete a task was used as a parameter to observe the difference between elite, intermediate and poorer teams [51]. For the sake of simplicity, the term *superior performers* is adopted in this review to describe experienced [51], elite [45], high-scoring [53], low-error

[54], skilled, good-quality [48] and any terminologies of similar sort.

- *Task-conditions*, differences in objective scores or any performance indicator could be used to learn the impact of various task conditions such as complexities i.e., high-intensity versus low-intensity phases of task [51, 55, 27].
- *Organisation*, differences in objective scores or any performance indicator due to differences in organisation of team. Influence of individual attitudes such as homogeneity, heterogeneity, evolution of ownership and leadership, and conflict on task performance are studied in the literature [48, 26, 56].
- *Familiarity*, performance differences that could be observed between familiar team members vs non-familiar team members. Some studies employ this approach to observe performance differences between pseudo-generated teams against original teams [27].
- *Information Exchange*, performance differences observed due to collective knowledge and skills of the team members, uniqueness towards exchanging information, collective problem solving and processing and deciphering information [2, 57].

In the event of multiple performance determining factors, cumulative scores representing the achievement of objectives are employed in some studies as task specific performance scores. These objective scores offer excellent platform to compare team performances.

IV. MEASURING TEAM MEMBERS' ACTIVITY

For understanding how PBR have contributed to team studies, this section briefly introduces different data acquisition methods observed in the literature. Focusing on the primary motive of this article, only methods that could have contributed to team performance related studies, in any closely related form are selected for the following discussion. Supplementary table.I briefly summarises all the techniques and their related measures discussed in this section. Cowley [58] discusses several sources to record the activity from individuals. However, techniques that are continuous, unobtrusive and non-confounding have been an integral part of team research studies as discussed below.

A. Physiological Observations

The physiological methods are primarily concerned with the internal activity from the human nervous system. Figure. 1 summarises the organisation of the human nervous system, key technologies for recording the activity from individuals, and various levels of measurements that are potentially useful for performance analysis of teams.

The human nervous system is composed of the central nervous system (CNS) and the peripheral nervous system (PNS). The activity of CNS can be captured using Electroencephalography (EEG) and Near-Infrared Spectroscopy (NIRS). Typically, wavelet or Fourier transformation could be used to decompose the temporal voltage information from EEG channels into six frequency components referred to as *delta* (1-3Hz), *theta* (4-7Hz), *alpha* (8-13Hz), *beta* (14-30Hz), *gamma* (31-50Hz)

and *mu*. In addition, these frequency components are used to compute different measures of cognitive states such as engagement EEG_E [59] and workload EEG_{WL} [59, 60, 55] of individuals. Furthermore, the EEG data can also be processed to account for eyes' blink referred to Eye Blink Rate (EBR) which is an estimate of blinks per minute [27]. From NIRS, levels of oxygen in haemoglobin in the form of oxy-Hb and deoxy-Hb can be recorded from individuals [47, 39].

The activity of PNS can be further sub-categorised into autonomic nervous system (ANS) and somatic nervous system (SNS) [8]. ANS has attracted special attention in team context because it reflects the activities due to interpersonal relationship [61]. The activity in ANS can be captured using Electrodermal activity (EDA), electrocardiogram (ECG) and respiration [26, 62, 8, 63, 24]. From EDA, measurements such as skin conductance level (SCL) and skin conductance response (SCR) (also referred to as galvanic skin response) are of interest in team studies [26, 24, 64]. From ECG, the intervals between R-peaks referred to as inter-beat interval (IBI) are used to measure heart rate (HR), Heart rate variability (HRV), and respiratory sinus arrhythmia (RSA) of individuals[17]. In addition, HRV-based measures such as (a) the high frequency component of HRV between 0.15Hz-0.40Hz (HRV-HF) [24], and (b) Statistical measures such as standard deviation of normal-normal (SDNN) [63, 62] and root mean of squared standard deviation (RMSSD) [62], are of interest in team studies.

The activity of SNS can be observed in HRV-HF, facial electromyography (EMG) and impedance cardiogram (ICG) [65]. From EMG, measurements such as orbicularis oculi (OO) and zygomaticus major (ZM) and currogator supercili (CS) are of interest in team studies [8, 40, 50]. From ICG, measures such as pre-ejection period (PEP) and left ventricular ejection time (LVET) have been used in team context [65].

B. Behavioural Observations

The behaviours of individuals can be observed in techniques such as communication recordings [2, 51, 66], eye-gaze tracking [58, 67] and body movements [48, 45] which convey valuable information in analysing and evaluating team performance.

Communication records can convey greater insights about shared cognition, team situational awareness, shared mental models and team-level interaction processes [10]. The team communication can be processed using either verbal or non-verbal approach [18]. Verbal contents include processing of audio transcripts to extract speech and linguistic features. This has been achieved using latent semantic analysis (LSA) [52], IBM speech to text services [68, 38] and Linguistic Inquiry and Word Count text analysis program [1]. Non-verbal content based measurements can include information about number of communication attempts such as button-clicks and acknowledgements to attempted communication [69], by extracting vocal tone frequencies [25, 30] and team interaction instances of attempted problem-solutions, general meetings, debriefing and brainstorming [37]. Further, coordination was also measured from communication such that for every information shared by a team member towards a specific member

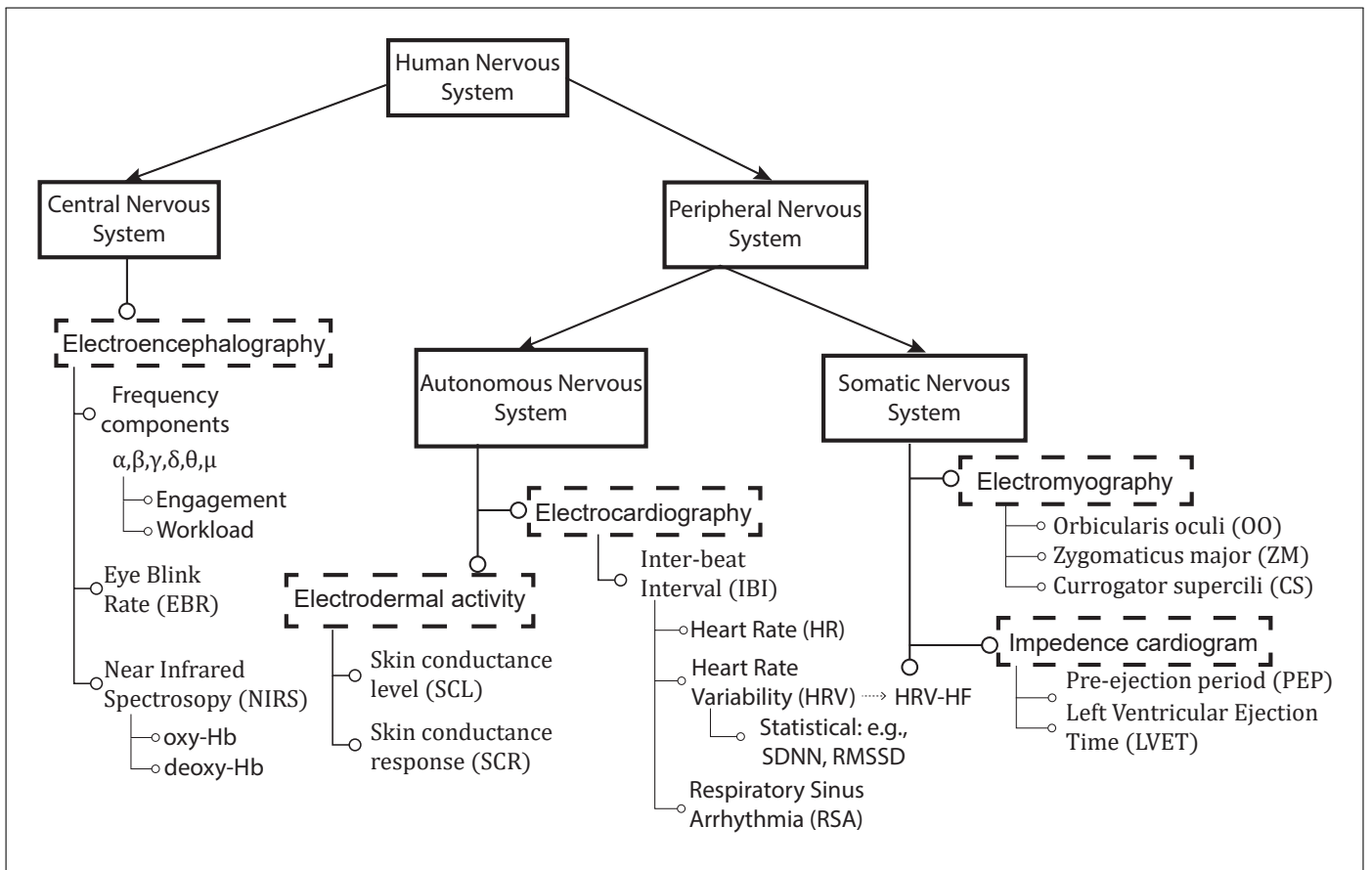


Fig. 1: A summary of physiological systems (dashed-boxes) for observing internal activity from individuals. Also shown are the potential measures at individual-level that are of importance for team level analysis

of the team who returned an acknowledging negotiation. This process was summarised using a cooperation index which is a ratio of number of communication attempts to cooperated acknowledgements. Complete details of non-content based measures that could be extracted from communication is discussed by Cooke *et al.* [70].

Eye movements are low-level manifestations of visual-cognitive interactions. Gaze data enables observing visual importances of individuals' providing valuable inferences underlying cognitive processes. Eye-gaze tracking offers a range of metrics such as fixations and saccades [71] to quantify the gaze behaviour of an individual. The measures can be used to observe the relationship about team members gaze behaviour and performance. Previously, gaze based measures such fixations around certain regions of interests, are employed to understand the impact of joint-activity on performance [67, 45].

Finally, team members actions has been analysed using accelerometers [48] or video recordings [72, 45, 66]. Physical bodily movements such as postural sway, hand coordination and face and head movements are also used to evaluate joint-activity in relation to performance [37, 38, 68, 45].

C. Multimodal Techniques

Aggregates of PBR employing measurement from more than one mode is referred to as multimodal measurements [58]. It is believed that the natural reactions of humans involves combinations of physiological activities and behavioural responses [8]. Multiple modalities thus can capture more than one response concurrently providing better representation of an individual's experience. Such an approach can aid better deciphering of intricate interpersonal physiological entrainment between the team members in relation to team performance [52, 27]. Multimodality can be achieved either by a combination of physiological and behavioural technique [52, 25], a combination of two or more physiological techniques [65, 27], or a combination of two or more behavioural techniques [73].

V. TEAM-LEVEL MEASUREMENTS

In a team context, the PBR at individual-level must be collectively analysed [17]. Such analyses are important for deriving quantitative and qualitative measures that could be used to evaluate, analyse and rationally comprehend the influence of constructs of IPO in relation to performance. The following metrics have been observed in the literature that could be used as indices of collective team response.

A. Compliance Metrics

Previously, compliance was quantified via neurodynamic synchrony (NS) which is a measure of similarity exhibited by team members' physiological measurements [8]. Note that, only physiological signals recorded from an identical source are deemed valid for deriving compliance metrics. Some of the compliance based measures are as follows:

1) *Directional Agreement (DA)*: represents a high-level estimate of compliance [24, 17]. When comparing physiological signals from two members of a team, DA represents the number of points that directionally matched either in the positive cycle or on the negative cycle in relation to their previous points. DA is expressed as percentage of points directionally matching from an entire data set.

2) *Signal Matching (SM)*: It is considered as a good measure of compliance [24, 17]. SM measures the area between two team members' physiological signals. Larger area indicates less similarity whereas smaller area indicated more similarity.

3) *Weighted Coherence (WC)*: It is a measure of similarity between two team members' physiological signals observed at specific frequency bands of interest [24, 17, 74].

4) *Instantaneous Derivative Matching (IDM)*: It represents a similarity measure between the curves of two team members' physiological signals [24]. The measure accounts for average number of differences between the derivatives observed at common reference points from the corresponding physiological signals [17].

5) *Regression*: Guastello demonstrated the employment of regression based models to measure the overall linear and nonlinear synchronisations between team members' autonomic arousals [64]. In this light, measures such as coefficient of determination, autoregression weight and synchronisation weight were derived from linear autoregression, and lyapunov exponent and dimension from autoregression weight of the exponential series. Furthermore, method such as cross-correlation was also used to measure the similarity between team members physiological signals [24, 53, 63, 47]. Besides, in the presence of multiple channel data (e.g., EEG data with N-channels), regression based methods have been employed to account for neurodynamic linkages i.e., active linkages in the brain regions of cooperating individuals [27, 55]. In this light, multivariate autoregression and partial directed coherence were used to derive the measurements, *inter-connectivity density* (ICD) and *inter-areas links density* (IALD) [27] also referred to as *strength density* and *clustering coefficient* [55], respectively.

6) *Cross Recurrence Quantification Analysis (CRQA)*: The CRQA is an alternative method to quantify nonlinear synchronisations between team members physiological or behavioural signals. It enables the concepts of recurrence quantification analysis and recurrence plots [75] to be extended for team based studies. The diagonal, horizontal and vertical profiles, and the density of points in the recurrence plot, respectively, are used to observe nonlinear properties such as, Determinism, Laminarity, Entropy and Rate of Recurrences. CRQA has been employed to observe the relationships such as joint activity and coordination between the team members [48, 67,

53, 49, 63]. Other versions of CRQA are multidimensional-RQA (MdRQA) [48, 38] also referred to as multivariate-RQA (MvRQA) [76] offer similar measurements but with an advantage of comparing multiple signals originating from two team individuals.

B. Temporal Neurodynamic Patterns of Teams

Physiological signals of team members have also been used to construct neurodynamic patterns. For instance, temporal measures of engagement EEG_E , workload EEG_{WL} and IBI from individual members of a team could be collectively represented to form team-level neurodynamic patterns [52, 51, 54, 77]. Such patterns represent instantaneous states of the team's engagement-level, workload-level and cognitive-load. Quantification of such patterns could be achieved using the concept of Shannon's Entropy where sliding windows of appropriate length are used to account for evolution of teams neurodynamic states [51]. For example, if a team experienced a variety of states during a window then the resulting Entropy would be high and *vice versa*. Such approaches enable observation of temporal variations in Entropy levels. Furthermore, assuming team individuals represent a complex dynamic systems the temporal Entropy series also reveals attractor basins. In addition, the neurodynamic patterns were also used to construct a 1-second lag transition matrix [51, 54]. The matrix reveal overall concentrations of different levels of teams engagement and workload.

C. Latent Semantic Analysis (LSA)

LSA can extract discriminating parameters from communication transcript such as semantic vector length and cosine-parameter that have been employed to observe differences between teams performance [52]. Semantic vector length is a measure of number of words either during a piece of discourse, phase of task, or during an entire task period. On the other hand, the LSA cosine-parameter measures the similarity between two transcripts or two pieces of discourse [78].

D. Epistemic Network Analysis (ENA)

ENA employs singular value decomposition at its kernel which is a type of unsupervised clustering. It constructs network such that nodes of the network represents key variables such as regions of interest and discussion topics, and the connection between nodes represent strength of interaction between the key variables. Such networks are used to analyse speech data, interaction between individuals in a group [79, 80] and also to understand gaze coupling between members of a team. Epistemic modelling is also advantageous in processing multimodal signals [37].

VI. EVALUATING RELATIONSHIPS BETWEEN TEAM PERFORMANCE AND CONSTRUCTS OF IPO

This section looks at how the analyses and techniques mentioned in the previous section transforms raw physiological and behavioural signals into team related dynamics. Also, the sections briefly unfolds the relationship between team dynamics, performance and the constructs of IPO framework.

A. Team Performance and Team Inputs

The inputs to teams i.e., the composition of teams and the structuring of task conditions significantly impacts team performance. Evaluating the relationship between team composition and task aids in unravelling the physiological operating dynamics of superior performers [35]. For instance, the transition matrix constructed using temporal EEG_E and EEG_{WL} states of the team demonstrated that teams composed of experienced members had more concentrations in the region of higher levels of team engagement and workload [51, 54]. Further, using hierarchical clustering algorithm on the LSA of cosine measured between all pairs of teams clearly clustered experienced teams and novice teams [51]. The sliding-window based entropy levels constructed from temporal EEG_E and EEG_{WL} demonstrated that experienced teams had relatively higher entropy levels indicating a greater neurodynamic flexible characteristics [52, 81]. A non-verbal team communication analysis by Cooke *et al.* [69] revealed that experienced teams exhibited a lower coordination to communication ratio.

The relationship between task structure and team performance could be used to identify conditions of task that impact performances. Indeed, LSA of team communication transcripts was employed to learn correlations between team performance scores and task conditions [52]. For example, vector lengths derived from LSA were used to identify the active phase of a task [52] such that, the phase of a task represented by smaller vector lengths was prominent during high active phase relative to other phases of the task.

In addition, the ICD indices could also be used to identify harder phases of a task [27, 55]. Any elevations in the regions of the brain's activity due to higher task complexity could be observed via the IALD analysis [27, 55]. Besides, the ICD also demonstrated a significantly higher neurodynamic linkage in the real teams during an elevated cooperation phase of the task relative to pseudo teams (synthetically generated from real teams' data) [27].

B. Team Performance and Team Processes

Understanding the relationships between performance and processes of teams can help explain the impact of qualities and emerging states on team performance.

In the context of cooperation and coordination, Wallot *et al.*'s work [48] enabled to observe the relationship between information sharing, behavioural coordination and performance. Accordingly, during an industrial production line simulation task, the homogeneity (egalitarian) condition, yielded better design of products relative to other conditions. Further, the quality of the product design inversely correlated with the determinism and laminarity (from MdRQA) between team individuals' hand movements (quantified using accelerometer readings). Similarly, FNIRS measure oxy-hb demonstrated that wavelet coherence between mixed gender pairs has significantly higher correlations under cooperating phases of a task [39].

The relationship between coordination and task performance could also be analysed using compliance metrics. As such, in a Multi-attribute task battery (MATB) environment, a lower

RMSE level during tracking was used as indicator of superior performance [24]. Montague *et al.* reports a significant positive correlation between the DA of team members SCR and performance [24]. Further, larger mean absolute deviation indicated poor resource management which correlated with lower SM indices (indicator of higher NS).

Compliance metrics are also employed to analyse the relationship between coordination and task conditions as a function of performance. The WC of HRV-HF captured the drop in performance under increased task complexity conditions [24] such that the WC resulting in higher averages under hard and low reliability condition relative to normal condition. In another custom developed tracking task, team performance was measured based on completion time, collision damage, RMS error (deviation from the centre path) as a direct objective of team coordination [82]. Henning *et al.* observed that WC of EDA and HR was related with drop in completion time. The WC of EDA, HR and respiration was also related with drop in RMS error. The CC of HR also demonstrated significant relationship with drop in completion time and RMS error. In their following study, CC of HR was able to predict performance changes as a function RMS error with significant negative correlations [83].

The neurodynamic linkages measured using ICD and IALD [27, 55] also corresponded to cooperation, and the corresponding analysis enabled observation of the relationship between levels of cooperation and performance.

In the context of joint activity, Tolston *et al.* [67] observed that the recurrence between the team members' gaze activity negatively correlated with performance scores. Video analysis has also been used to observe coordination as a function of team composition [45]. A video analysis of laparoscopic surgery task revealed that teams composed of elite surgeons exhibited low de-synchronisations relative to the other composition of teams.

Communication has been employed to understand the relationship between aggregates of team cognition and performance [2, 70]. Relationship between team performance and experience was explored by Cooke *et al.* [2]. Accordingly, it was observed that team performance improved as teams gained experience, which was resultant of improvements in team interaction (process behaviours such as coordination and communication). This indicates that the development of shared mental models improved performance [2]. Further, it was observed that the teams which resumed task after shorter breaks performed better than teams that took longer breaks between tasks. This decline performance was observed due to loss of skill by individuals rather than team interaction (interaction-based measures) [70]. This shows that degradation of skill was significant compared to shared mental models.

C. Team Performance Outcome and Its Implications

The implications of team performance outcomes is essential for effective evaluation of underlying team inputs and processes. Previously, this approach has been used to understand shared decision making due to performance feedback. In a production line task involving crafting of origami sail boats

[49, 50], correlation between NS and performance developments were reported. During a negative performance development, teams took a call for changing the strategy. When teams were presented with new strategy and post adoption of the strategy, the NS between ZM and SCR of team members' was observed to be low. On the other hand, NS of SCR increased in teams which opted to continue without adopting the presented strategy.

This context of analysis can also be applied in the prediction of critical performance outcomes. As such, in a programming scenario, where performance of teams was decided via pass or fail outcome of the task, SCR was able to distinguish the effect of outcome in advance. During a passed outcome, the dyad demonstrated significantly higher SCR levels compared to failed outcome [26].

Indeed, this approach was employed to observe the effect of feedback on cooperation. Using NIRS measures of oxy-Hb level, Balconi *et al.* observed a drop in inter-brain synchronicity following an induced negative feedback [47]. Also, Cheng *et al.* observed a positive correlation between task performance and synchronicity [39]. Such activities in specific regions of brain can be used as markers of cooperation [47, 39, 84].

The performance outcome can also be used in objective identification of skill composition or characteristics of teams. It was observed that as novice teams gained experience via subsequent repeating runs through the task, there was a gradual rise in the mean entropy [51] which could be used as indicators of a teams performance evolution.

VII. MULTIMODALITY IN THE CONTEXT OF TEAMS

The team inputs, processes, states, qualities and outcomes are codependent entities. In an ecological context, combination of these entities interplay defining the natural behaviour of humans [8]. In this light, multiple-PBRs have been previously employed to study human emotions to understand collaboration, cooperation and coordination [25, 27, 26, 85]. Supplementary table. II provides a summary of methods contributing to multimodality in team context.

Individual emotions can be associated with multiple interaction processes which can be effectively evaluated using multimodality. For example, the 'threat' is an emotional state in humans that can elevate vasoconstriction and reduce in cardiac functioning which is directly responsible for inducing stress [25]. The heightened stress levels can diminish cognitive functioning (the ability to think and response) causing ineffective communication. This phenomenon was observed and validated by Neubauer *et al.* using speech and cardiac data [25]. Vocal-tone variations from speech data and total peripheral resistance from ECG measurements were used in their definition of a multimodal index [25]. The index was used to understand the impact of stress leading to communication response.

Analogously, the investigation of *cross-level effects* by Gorman *et al.* observed simultaneous developmental changes in cognitive and behavioural functioning due to team interactions [52]. EEG and communication data were analysed to understand entrainment between team members as they progressed through a task. Such analysis enables better task related performance evaluation that result from team-level entrainment, for

example, in understanding leader driven theories [86, 56, 87] where speech from the leader induces motivation that affects team-processes eventually reflecting in enhanced performance [52].

Multimodal approaches have also been favoured in enhancing team training outcomes and assessments [52, 36]. For example, concurrent data from eye-tracking and EEG were collectively analysed to observe the impact of cohesiveness and expertise in a cybersecurity setting [36]. Collective analysis using additional modality also provide advantage by potentially substituting for any gaps that might be suffered through a single mode [88, 25]. Toppi *et al.* observed that HR and EBR, which are regulated by CNS activity demonstrated strong correlations with different task conditions [27]. The correlations also significantly varied depending on operator roles and responsibilities. Such analyses can be employed to understand relationship between team inputs *i.e.*, team composition and task conditions.

Multimodal research studies enable identification of substitute sources that offer cost-effective and better alternative solutions. For example, by comparing concurrent measures of EDA and ECG, Ahonen *et al.* was able to observe that SCR (from EDA) offered better and faster method to convey information originating from ANS activations when compared with HRV features (from ECG) [26].

Multimodality aids to explore the impact of certain stimuli on team outcomes. For example, (a) Murray *et al.* explored feature vectors comprising of features from both linguistic and acoustic speech modulations to observe relationship with task performance [73], and (b) Neubauer *et al.* explored the relationship between acoustic voice features and physiological states [25]. In the context of studies mentioned in (a) and (b), multimodal analysis can be used to link both the physiological and behavioural responses to better explain task outcomes.

Overall, it can be established that multimodal approaches have the capability to promote significant and distinct observations, where a single modality would not have comprehensively explained the scenario [37, 38].

A. Data Fusion in Multimodal Studies

In addition to the employment of multiple modalities, a collective analytical index which mathematically and rationally unifies data from multiple modalities is very important. Only a few studies have attempted data fusion from multiple modalities or collective analysis in some form [52, 65, 37, 38, 49, 85]. The method demonstrated by Eloy *et al.* employs measurements from three different modalities [38]. In this method, for any member the three PBR signals were first averaged. Then, the averaged signals from all team members were analysed using MdrQA to estimate unified percentage recurrence. Echeverria *et al.* employs a multimodal matrix mechanism to generate an interactivity network to observe teams' performances at various scenarios [37]. Monster *et al.*'s method employs joint activity measurements derived from ZM and SCR of all the teams [49], where, principal component analysis was later employed to observe distinct groups. Finally, the procedure employed by Gorman *et al.* analyses each modality

individually, however, discussions link observations from EEG and speech data analysis, explaining cross-level effects [52].

VIII. SIMULATION ENVIRONMENTS

Simulation environment also referred to as synthetic task environment (STE) is an integral component in contextualising team research studies [10, 89, 90]. They offer a control, cost-effective and safer working platform for team individuals to interact and share a visual space. Such platforms enable easy modifications of task conditions, regenerate and adjust task elements according to required activity [91] thereby, enabling enhanced training outcomes and economical team training methods [89]. Ideally, a platform that can emulate natural settings is a preferred choice. However, team studies have employed environments ranging from a simple board game [92] to multiplayer computer games(e.g., Bomber man, Ibb & Obb, Tetris and Hedgewars) [40, 93, 94, 33, 17, 95, 53] and role play [7] which offer choices to observe coordination and cooperation. To observe compliance in behavioural responses against team performance, physical tasks such as origami construction[50, 49], lego building [48], or even head movements [38] have also been observed in the literature.

In recent years, simulated environments have attracted attention in Defence and civil sectors as an efficient and effective mechanism for training [96]. STEs that emulate real world scenarios offer the perfect testing grounds for teams. This includes ecological settings such as flight simulator [27, 29] and hands on surgical setup [37, 97, 45]. Tactical environments that offer command and control operations are typical choices for team cognition study. Example of tactical environments include submarine piloting and navigation [52, 51], unmanned aerial vehicle-STE [69, 70], research environment for supervisory control of heterogeneous unmanned [67] and many similar forms can be observed in [98, 89, 99]. In addition, Live virtual and constructive environments are fully integrated simulators specifically employed by Defence Forces in many countries [91, 100, 101, 102]. It integrates the concepts of augmented reality and virtual reality to provide near realistic air-combat and land-combat training experience to defence personnel.

Non-tactical approaches offering controlled environments to study cognition included choices such as Multi-Attribute Task Battery [55, 24], process control synthetic environments [65] and customised cooperative task [84, 39]. These platforms offer integrated system equipped with variety of training benefits such as lower costs, safety and, improved training outcomes.

A. Current Limitations of Synthetic Environments

Recently, Lawson *et al.* reviewed 57 commercially available and licenced computerized task environments for assessment of military team performance [89]. The review, however, does not focus on the convenience of employing such environments with PBR devices. It is very important to ensure that the PBR do not interfere participants' commitment with the task. For example, finger-tip connectable devices may interfere when using mouse clicking or keyboard operation. Similarly,

tasks projected on multiple screens result in constant neck movement interfering with desktop based eye-tracking devices. Further, the primary motivation of commercial games was to meet entertainment demands, hence, may not be effective for research purposes. Such task environments may not offer acquisition of data related to task dynamics and user experience that can effectively complement PB recordings. The lack of which impedes comprehending temporal correlations between task dynamics and PB recordings. In addition, majority of game environments can only engage two persons to small groups, some challenges could be complex for general public, or may not be suitable for all age groups. Such drawbacks limit efficient addressing of team related research questions.

IX. SUBJECTIVE RATINGS

Subjective rating are questionnaires which are used in studies to obtain feedback from participants [40, 83]. Such an approach has been traditionally used where individuals rate their experience concerning task difficulty, emotion, and personality-traits [103, 94, 50]. NASA Task Load Index (NASA-TLX) is one of the commonly employed subjective rating techniques to understand the task workload experienced by members of a team [98, 97, 55]. Later, a team-level index can be obtained by globally averaging the scores from individual's subjective ratings.

Further, in the light of "human emotions", the Dundee Stress State Questionnaire [1], the Trait Anxiety-Taylor Manifest Anxiety Scale [64], the State-Trait Anger Expression Inventory-2[103, 18], the Profile of Mood States [104], the Positive and Negative Affect Scale [40], the Self-Assessment Manikin (SAM) [105, 40]. In addition, questionnaires such as Emotional Intelligence [64], Coping flexibility[64] and Empathy Assessment Index [64] were found in the literature that specifically reflect human emotions.

To assess the competitiveness and cooperativeness in computer game environments or social gaming platforms, questionnaires such as the Competitive and Cooperative Presence in Gaming Questionnaire[93], the Social Presence in Gaming Questionnaire [40], and the checklist for trust and team affect questionnaire [99] have been employed.

In the light of team cognition, task specific questionnaires, situational Awareness Rating Technique [98, 106], Crew Awareness Rating Technique [107, 106] and Situation Awareness Alobal Assessment Technique [106] were observed in the literature.

X. DISCUSSION AND FUTURE RECOMMENDATIONS

Salas emphasised the need for modern technologies to promote extensive and good quality team research [10]. Current advancements in wearable sensor technologies could boost multimodal-based team research [90]. Several low-cost, low-powered and smart devices are currently flooding the consumer market enabling better integration of off-shelf wearable and portable PB sensors and technologies into experiments [108]. Observing this trend, it can be realised that PBR technologies have been very well advancing, relatively however, the processing and analysing of the data has slowly progressed,

more so, methods for collective team related analysis. Some of the constraints that have been constantly highlighted by researchers are (a) computationally demand (b) expensive setup [73] (c) identification of right combination of recording systems [26] and (d) effective fusion of data from multiple modalities are sparsely observed in the literature. Further, following are some of the key areas that studies concerning team performance analysis could focus in the future research.

A. Team Cognition Based Performance Evaluation

Team cognition has been highly valued among work teams [10, 109], and interactive dynamics strongly reflects these aspects. This review reveals that team cognition has been mostly studied in terms of interactivity using communication and eye-tracking [2, 109]. Literature mentions physiological measures such as cardiac [62, 94, 90] as markers for analysing team cognitive activity, however, markers informing team cognition aggregates such as shared mental models needs empirical reinforcements. For example, comparing studies [25] and [2], where both employ some form of measurements derived from communication. Towards generalisation and in natural scenarios, the ITC study [2] would also benefit with the inclusion of additional modes such that, any interpersonal and emotional factors that may have affected or promoted team performance [47] could be accounted via an additional modality.

Literature mentions three primary means to observe cognition in teams: interactions, team knowledge and mental models, and team situational awareness (SA) [110, 90]. In this light, real-time concurrent measures from EEG, eye-tracking and communication transcripts were emphasised in the literature [110]. Further, Cooke *et al.* states that studies must link performance outcomes with emergent states such as cognition rather than linking to physiological measures at individual level [2]. Thus, multimodality could be key for efficient deciphering of attributes underlying team performance.

B. Analysis Beyond Dyads

Majority of the team studies observed in the literature are mostly concerned with dyads [111, 24, 109] where complementary emergent states are inherent. In contrast, limited studies have explored teams comprising of a minimum three persons or larger. Amongst these, physiological studies investigating neurodynamic organisation in six-member teams [51, 52], NS analysis in four-member teams [17] and joint activity computed for three-member teams were observed in the team performance analysis literature [49]. Further, in some studies though teams larger than dyads were considered, the PC and joint activity was computed between any two members (hereafter referred to as sub-dyads) of a team [17, 49]. Later, the physiological interactivity that was measured across all possible sub-dyads of a team was globally averaged to compute a team-level interactivity score.

Future team experiments could engage three or more members in team to move team-based analysis beyond dyads. Ideally, studies must focus on concurrently analysing multiple team members data that could be represented using a unified index

such as the neurodynamic organisation [51, 52], or interactivity index using multimodal features by Eloy *et al.* [38]. Further, methods employed in sub-dyadic approach [17, 49], could have better impact via more empirical evidences. Such advancements are necessary for standardising and generalisation any innovative methods across team study community. Addressing such challenges fosters newer analytical methods for efficient deciphering of attributes underlying team performance.

C. Need for Computation Improvement

Transferability of computation methods across the broad ramifications of team performance can enable newer data fusion methods. Further, empirical evidences concerning data processing methods are relatively less observed in the literature. For example, nonlinear dynamics in team analyses were observed to be less compared to the use of linear statistics. In addition, the employment of artificial neural networks is still basic. Only one study was observed to use artificial neural network based future outcome prediction [29]. Our review also show that temporal prediction capabilities of recurrent neural networks have not been observed in team settings. There is a need for maximising the sample count using boot-strapping [62] and pseudo team data generation [27] from small collection. Such techniques can encourage the employment and development of deep neural networks based team research studies.

D. Experimental setup

The experimental setup should support continuous acquisition of PBRs, and concurrently from all the members of team. There is a need for task environment that can constantly engage members of team as opposed to the nature of distributed task environment [64]. Such environments can induce interactions enabling evaluation of team performance in relation to adaptability, emergence of leadership and coordination [112]. Task environments that can engage general public improve sample count as well enable exploring a diverse pool of skills. The environment must offer engagement of more than two participants to advance team research beyond coupled dynamics. The stimulus or testing tasks built into the environment must be strong enough to test the targeted individual behaviours and performances [48]. The environment must offer compactness, portability, easy configurability and cross-domain compliant [96]. It must be capable of testing cooperation and coordination, and provide qualitative or quantitative feedback of team performance. Such feedback mechanism reflect behaviour related to cognitive activity [47]. In addition, task-event data that can temporally unfold dynamics of the task such as mouse clicks, movements, key-element usage and key-logs, must be accessible from the environment [19].

This review revealed that majority of team studies have experimented with physiological systems compared to behavioural systems. In this light, future research could focus on a combination of physiological and behavioural system targeting greater flexibility in team studies. Specifically the employment of eye-tracking and communication is encouraged to better

explain team cognitive aggregates such as share mental models and team situational awareness [110]. Finally, well targeted subjective ratings must be included as a means to collect direct feedback from participants.

XI. CONCLUSION

Literature concerning performance analysis of ER&SC teams using physiological and behavioural approach are not exclusive, however, are critical for many team-based research. Furthermore, the literature is significant for employment of multimodality in ER&SC team research. The literature pool suggests that the advancement of multimodality in general demands for a strong multidisciplinary approach. As such, this review briefly summarised single modality ER&SC team performance studies and methods employed for data processing and how contemporary experiments could be shaped. The need for dedicated STE and sensor integration realises that team research can benefit from the involvement of broader research communities. Such developments can promote advanced nonlinear neurodynamic analysis of teams and newer mathematical models such as data fusion techniques. Overall, contributing to enriching the literature pool, empirical evidences, and subjective performance quantification and analysis of ER&SC teams.

REFERENCES

- [1] Catherine Neubauer, Joshua Woolley, Peter Khooshabeh, and Stefan Scherer. Getting to know you: a multimodal investigation of team behavior and resilience to stress. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 193–200, 2016.
- [2] Nancy J Cooke, Jamie C Gorman, Christopher W Myers, and Jasmine L Duran. Interactive team cognition. *Cognitive science*, 37(2):255–285, 2013.
- [3] David P Baker, Rachel Day, and Eduardo Salas. Teamwork as an essential component of high-reliability organizations. *Health services research*, 41(4p2):1576–1598, 2006.
- [4] Eduardo Salas, Kevin C Stagl, C Shawn Burke, and Gerald F Goodwin. Fostering team effectiveness in organizations: Toward an integrative theoretical framework. In *Nebraska symposium on motivation*, volume 52, page 185, 2007.
- [5] Jamie C Gorman, Nancy J Cooke, and Polemnia G Amazeen. Training adaptive teams. *Human Factors*, 52(2):295–307, 2010.
- [6] Jamie C Gorman, Polemnia G Amazeen, and Nancy J Cooke. Team coordination dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, 14(3):265, 2010.
- [7] Selma van der Haar, Mieke Koeslag-Kreunen, Eline Euwe, and Mien Segers. Team leader structuring for team effectiveness and team learning in command-and-control teams. *Small group research*, 48(2):215–248, 2017.
- [8] Sadaf Kazi, Salar Khaleghzadegan, Julie V Dinh, Mark J Shelhamer, Adam Sapirstein, Lee A Goeddel, Nnenna O Chime, Eduardo Salas, and Michael A Rosen. Team physiological dynamics: A critical review. *Human factors*, page 0018720819874160, 2019.
- [9] Michael A Rosen, Aaron S Dietz, Ting Yang, Carey E Priebe, and Peter J Pronovost. An integrative framework for sensor-based measurement of teamwork in health-care. *Journal of the American Medical Informatics Association*, 22(1):11–18, 2015.
- [10] Eduardo Salas, Nancy J Cooke, and Michael A Rosen. On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, 50(3):540–547, 2008.
- [11] Gregory J Funke, Benjamin A Knott, Eduardo Salas, Davin Pavlas, and Adam J Strang. Conceptualization and measurement of team workload: A critical need. *Human Factors*, 54(1):36–51, 2012.
- [12] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. A temporally based framework and taxonomy of team processes. *Academy of management review*, 26(3):356–376, 2001.
- [13] JR Hackman. The design of work teams. handbook of organizational behavior. edited by: Lorsch jw. 1987.
- [14] Günther Knoblich and Natalie Sebanz. The social nature of perception and action. *Current Directions in Psychological Science*, 15(3):99–104, 2006.
- [15] Norman Warner, Michael Letsky, and Michael Cowen. Cognitive model of team collaboration: Macro-cognitive focus. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 269–273. Sage Publications Sage CA: Los Angeles, CA, 2005.
- [16] Herbert H Clark. Coordinating with each other in a material world. *Discourse studies*, 7(4-5):507–525, 2005.
- [17] Amanda N Elkins, Eric R Muth, Adam W Hoover, Alexander D Walker, Thomas L Carpenter, and Fred S Switzer. Physiological compliance and team performance. *Applied ergonomics*, 40(6):997–1003, 2009.
- [18] Barthélémy Bertrand, Jean-Noël Evain, Juliette Piot, Rémi Wolf, Pierre-Marie Bertrand, Vincent Louys, Hugo Terrisse, Jean-Luc Bosson, Pierre Albaladejo, and Julien Picard. Positive communication behaviour during handover and team-based clinical performance in critical situations: a simulation randomised controlled trial. *British Journal of Anaesthesia*, 126(4):854–861, 2021. Verbal and nonverbal.
- [19] Claire La Fleur, Blaine Hoffman, C Benjamin Gibson, and Norbou Buchler. Team performance in a series of regional and national us cybersecurity defense competitions: Generalizable effects of training and functional role specialization. *Computers & Security*, 104:102229, 2021.
- [20] Hsing-Yuan Liu. The relationship between swift trust and interaction behaviors on interdisciplinary and non-interdisciplinary teams in nursing education. *Nurse Education in Practice*, 51:102977, 2021.
- [21] Gregory J Funke and Scott M Galster. The effects of cognitive processing load and collaboration technology

- on team performance in a simulated command and control environment. *International Journal of Industrial Ergonomics*, 39(3):541–547, 2009.
- [22] Eduardo Salas, Denise L Reyes, and Amanda L Woods. The assessment of team performance: Observations and needs. In *Innovative assessment of collaboration*, pages 21–36. Springer, 2017.
- [23] David Dunning, Kerri Johnson, Joyce Ehrlinger, and Justin Kruger. Why people fail to recognize their own incompetence. *Current directions in psychological science*, 12(3):83–87, 2003.
- [24] Enid Montague, Jie Xu, and Erin Chiou. Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. *IEEE Transactions on Human-Machine Systems*, 44(5):614–624, 2014.
- [25] Catherine Neubauer, Mathieu Chollet, Sharon Mozgai, Mark Dennison, Peter Khooshabeh, and Stefan Scherer. The relationship between task-induced stress, vocal changes, and physiological state during a dyadic team task. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 426–432, 2017.
- [26] Lauri Ahonen, Benjamin Ultan Cowley, Arto Hellas, and Kai Puolamäki. Biosignals reflect pair-dynamics in collaborative work: Eda and ecg study of pair-programming in a classroom environment. *Scientific reports*, 8(1):3138, 2018.
- [27] Jlenia Toppi, Gianluca Borghini, Manuela Petti, Eric J He, Vittorio De Giusti, Bin He, Laura Astolfi, and Fabio Babiloni. Investigating cooperative behavior in ecological settings: an eeg hyperscanning study. *PLoS one*, 11(4):e0154236, 2016.
- [28] Chad L Stephens, Angela R Harrivel, Alan T Pope, and Lawrence J Prinzel III. System and method for training of state-classifiers, January 29 2019. US Patent 10,192,173.
- [29] Angela R Harrivel, Chad L Stephens, Robert J Milletich, Christina M Heinich, Mary Carolyn Last, Nicholas J Napoli, Nijo Abraham, Lance J Prinzel, Mark A Motter, and Alan T Pope. Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing. In *AIAA Information Systems-AIAA Infotech@ Aerospace*, page 1135. 2017.
- [30] Jennifer K Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Alevan. Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology*, 51(5):1527–1547, 2020.
- [31] Rakesh Veerabhadrappe, Masood Ul Hassan, James Zhang, and Asim Bhatti. Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting. *Frontiers in Systems Neuroscience*, 14:34, 2020.
- [32] RH Stevens, Trysha Galloway, Chris Berka, and Adrienne Behneman. A neurophysiologic approach for studying team cognition. In *Interservice/Industry Training Simulation and Education Conference (IITSEC), Paper*, number 10135, 2010.
- [33] Guillaume Chanel, J Matias Kivikangas, and Niklas Ravaja. Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interacting with Computers*, 24(4):306–316, 2012.
- [34] Semin Park, Travis J Grosser, Adam A Roebuck, and John E Mathieu. Understanding work teams from a network perspective: A review and future research directions. *Journal of Management*, page 0149206320901573, 2020.
- [35] Suzanne T Bell. Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of applied psychology*, 92(3):595, 2007.
- [36] Austin Silva, Glory Emmanuel, Jonathan T McClain, Laura Matzen, and Chris Forsythe. Measuring expert and novice performance within computer security incident response teams. In *International Conference on Augmented Cognition*, pages 144–152. Springer, 2015.
- [37] Vanessa Echeverria, Roberto Martinez-Maldonado, and Simon Buckingham Shum. Towards collaboration translucence: Giving meaning to multimodal group data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2019.
- [38] Lucca Eloy, Angela EB Stewart, Mary Jean Amon, Caroline Reinhardt, Amanda Michaels, Chen Sun, Valerie Shute, Nicholas D Duran, and Sidney D’Mello. Modeling team-level multimodal dynamics during multiparty collaboration. In *2019 International Conference on Multimodal Interaction*, pages 244–258, 2019.
- [39] Xiaojun Cheng, Xianchun Li, and Yi Hu. Synchronous brain activity during cooperative exchange depends on gender of partner: A fnirs-based hyperscanning study. *Human brain mapping*, 36(6):2039–2048, 2015.
- [40] J Matias Kivikangas, Jari Kätsyri, Simo Järvelä, and Niklas Ravaja. Gender differences in emotional responses to cooperative and competitive game play. *PLoS one*, 9(7):e100318, 2014.
- [41] Jeffery A LePine, Ronald F Piccolo, Christine L Jackson, John E Mathieu, and Jessica R Saul. A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61(2):273–307, 2008.
- [42] Karen A Jehn, Lindred Greer, Sheen Levine, and Gabriel Szulanski. The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Decision and Negotiation*, 17(6):465–495, 2008.
- [43] Hayden JR Woodley, Matthew JW McLarnon, and Thomas A O’Neill. The emergence of group potency and its implications for team effectiveness. *Frontiers in psychology*, 10:992, 2019.
- [44] Stanley C Collyer and Gerald S Malecki. Tactical decision making under stress: History and overview. 1998.
- [45] Wenjing He and Bin Zheng. Collaborative performance in laparoscopic teams: behavioral evidences from simulation. *Surgical endoscopy*, 30(10):4569–4574, 2016.
- [46] Oleg Špakov, Howell Istance, Kari-Jouko Rähkä, Tiia

- Viitanen, and Harri Siirtola. Eye gaze and head gaze in collaborative games. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pages 1–9, 2019.
- [47] Michela Balconi, Laura Gatti, and Maria Elide Vanutelli. When cooperation goes wrong: brain and behavioural correlates of ineffective joint strategies in dyads. *International Journal of Neuroscience*, 128(2):155–166, 2018. NIRS.
- [48] Sebastian Wallot, Panagiotis Mitkidis, John J McGraw, and Andreas Roepstorff. Beyond synchrony: joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PloS one*, 11(12):e0168306, 2016.
- [49] Dan Mønster, Dorthe Døjbak Håkonsson, Jacob Kjær Eskildsen, and Sebastian Wallot. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & behavior*, 156:24–34, 2016.
- [50] Dorthe Døjbak Håkonsson, Jacob Kjær Eskildsen, Linda Argote, Dan Mønster, Richard M Burton, and Børge Obel. Exploration versus exploitation: Emotions and performance as antecedents and consequences of team decisions. *Strategic Management Journal*, 37(6):985–1001, 2016.
- [51] Ronald Stevens, Trysha Galloway, Peter Wang, Chris Berka, Veasna Tan, Thomas Wohlgemuth, Jerry Lamb, and Robert Buckles. Modeling the neurodynamic complexity of submarine navigation teams. *Computational and mathematical organization theory*, 19(3):346–369, 2013.
- [52] Jamie C Gorman, Melanie J Martin, Terri A Dunbar, Ronald H Stevens, Trysha L Galloway, Polemnia G Amazeen, and Aaron D Likens. Cross-level effects between neurophysiology and communication during team training. *Human factors*, 58(1):181–199, 2016.
- [53] Adam J Strang, Gregory J Funke, Sheldon M Russell, Allen W Dukes, and Matthew S Middendorf. Physio-behavioral coupling in a cooperative team task: Contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):145, 2014.
- [54] Ronald H Stevens and Trysha L Galloway. Toward a quantitative description of the neurodynamic organizations of teams. *Social Neuroscience*, 9(2):160–173, 2014.
- [55] Nicolina Sciaraffa, Gianluca Borghini, Pietro Aricò, Gianluca Di Flumeri, Alfredo Colosimo, Anastasios Bezerianos, Nitish Thakor, and Fabio Babiloni. Brain interaction during cooperation: Evaluating local properties of multiple-brain network. *Brain sciences*, 7(7):90, 2017.
- [56] Eduardo Salas, Dana E Sims, and C Shawn Burke. Is there a “big five” in teamwork? *Small group research*, 36(5):555–599, 2005. Leadership skills.
- [57] Heidi Kristina Westli, Bjørn Helge Johnsen, Jarle Eid, Ingvil Rasten, and Guttorm Brattebø. Teamwork skills, shared mental models, and performance in simulated trauma teams: an independent group design. *Scandinavian journal of trauma, resuscitation and emergency medicine*, 18(1):1–8, 2010.
- [58] Benjamin Cowley, Marco Filetti, Kristian Lukander, Jari Tornainen, Andreas Henelius, Lauri Ahonen, Oswald Barral, Ilkka Kosunen, Teppo Valtonen, Minna Huotilainen, et al. The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction. *Foundations and Trends® in Human–Computer Interaction*, 9(3-4):151–308, 2016.
- [59] T. McMahan, I. Parberry, and T. Parsons. Evaluating electroencephalography engagement indices during video game play. In *FDG*, 2015.
- [60] Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, Paola Lanzi, Simone Pozzi, Valeria Vignali, Claudio Lantieri, Arianna Bichicchi, Andrea Simone, and Fabio Babiloni. Eeg-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Frontiers in Human Neuroscience*, 12:509, 2018.
- [61] Richard V Palumbo, Marisa E Marraccini, Lisa L Weyandt, Oliver Wilder-Smith, Heather A McGee, Siwei Liu, and Matthew S Goodwin. Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, 21(2):99–141, 2017.
- [62] Lauri Ahonen, Benjamin Cowley, Jari Tornainen, Antti Ukkonen, Arto Vihavainen, and Kai Puolamäki. Cognitive collaboration found in cardiac physiology: Study in classroom environment. *PloS one*, 11(7):e0159178, 2016.
- [63] Rakesh Veerabhadrapa, Imali T. Hettiarachchi, and Asim Bhatti. Using recurrence quantification analysis to quantify the physiological synchrony in dyadic ecg data. In *2021 IEEE International Systems Conference (SysCon)*, pages 1–8, 2021.
- [64] Stephen J Guastello. Physiological synchronization in a vigilance dual task. *Nonlinear dynamics, psychology, and life sciences*, 20(1):49–80, 2016.
- [65] Alexander D Walker, Eric R Muth, Fred S Switzer III, and Patrick J Rosopa. Predicting team performance in a dynamic environment: A team psychophysiological approach to measuring cognitive readiness. *Journal of Cognitive Engineering and Decision Making*, 7(1):69–82, 2013.
- [66] Ashish Malviya, Glenn A Fink, Landon Segó, and Barbara Endicott-Popovsky. Situational awareness as a measure of performance in cyber security collaborative work. In *2011 Eighth International Conference on Information Technology: New Generations*, pages 937–942. IEEE.
- [67] Michael T Tolston, Adam J Strang, Gregory J Funke, Brent Miller, Rebecca Brown, and Lauren Menke. Evaluating the relationship between team performance and joint attention with longitudinal multivariate mixed models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, pages 740–744. SAGE Publications Sage CA: Los Angeles, CA, 2016.

- [68] Hana Vrzakova, Mary Jean Amon, Angela Stewart, Nicholas D Duran, and Sidney K D’Mello. Focused or stuck together: multimodal patterns reveal triads’ performance in collaborative problem solving. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 295–304, 2020.
- [69] Nancy J Cooke, Jamie C Gorman, Jasmine L Duran, and Amanda R Taylor. Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied*, 13(3):146, 2007.
- [70] Nancy J Cooke and Jamie C Gorman. Interaction-based measures of cognitive systems. *Journal of cognitive engineering and decision making*, 3(1):27–46, 2009.
- [71] Stephanie Brams, Gal Ziv, Oron Levin, Jochim Spitz, Johan Wagemans, A Mark Williams, and Werner F Helsen. The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological bulletin*, 145(10):980, 2019.
- [72] Chao Zhang, Connie Miller, Kathleen Volkman, Jane Meza, and Katherine Jones. Evaluation of the team performance observation tool with targeted behavioral markers in simulation-based interprofessional education. *Journal of Interprofessional Care*, 29(3):202–208, 2015.
- [73] Gabriel Murray and Catharine Oertel. Predicting group performance in task-based interaction. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 14–20, 2018.
- [74] Robert Arthur Henning, Wolfram Boucsein, and Monica Claudia Gil. Social–physiological compliance as a determinant of team performance. *International Journal of Psychophysiology*, 40(3):221–232, 2001.
- [75] Norbert Marwan, M Carmen Romano, Marco Thiel, and Jürgen Kurths. Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6):237–329, 2007.
- [76] Panagiotis Mitkidis, John J McGraw, Andreas Roepstorff, and Sebastian Wallot. Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiology & behavior*, 149:101–106, 2015.
- [77] Roger D Dias, Marco A Zenati, Ronald Stevens, Jennifer M Gabany, and Steven J Yule. Physiological synchronization and entropy as measures of team cognitive load. *Journal of biomedical informatics*, 96:103250, 2019.
- [78] Nancy J Cooke. Using team communication to understand team cognition in distributed vs. co-located mission environments. *Applied Psychology*, 20051212:044.
- [79] David Williamson Shaffer, Wesley Collier, and Andrew R Ruis. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3):9–45, 2016.
- [80] Zachari Swiecki and David W Shaffer. *Toward a taxonomy of team performance visualization tools*. International Society of the Learning Sciences, Inc.[ISLS], 2018. ENA for team performance visualisation.
- [81] Frank E Ritter, William G Kennedy, and Bradley J Best. The best papers from brims 2011: models of users and teams interacting. *Computational and Mathematical Organization Theory*, 19(3):283–287, 2013.
- [82] Robert A Henning and Kristopher T Korbelak. Social-psychophysiological compliance as a predictor of future team performance. *Psychologia*, 48(2):84–92, 2005.
- [83] RA Henning, AG Armstead, and JK Ferris. Social psychophysiological compliance in a four-person research team. *Applied ergonomics*, 40(6):1004–1010, 2009.
- [84] Xu Cui, Daniel M Bryant, and Allan L Reiss. Nirs-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, 59(3):2430–2437, 2012.
- [85] Guillaume Chanel, Mireille Bétrancourt, Thierry Pun, Donato Cereghetti, and Gaëlle Molinari. Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 116–122. IEEE, 2013.
- [86] Paul Barach and Matthew B Weinger. Trauma team performance. *Trauma*, pages 145–158, 2007.
- [87] Jaap Schaveling, Saskia Blaauw, and Kees van Montfort. Predictors of group performance in a police criminal investigation department: the role of gender homogeneity, leadership and team characteristics. *Journal of Police and Criminal Psychology*, 32(4):358–368, 2017.
- [88] Joyce E Bono, Theresa M Glomb, Winny Shen, Eugene Kim, and Amanda J Koch. Building positive resources: Effects of positive events and positive reflection on work stress and health. *Academy of Management Journal*, 56(6):1601–1627, 2013.
- [89] Ben D Lawson, Thomas W Britt, Amanda M Kelley, Jeremy R Athy, and Shauna M Legan. Computerized tests of team performance and crew coordination suitable for military/aviation settings. *Aerospace Medicine and Human Performance*, 88(8):722–729, 2017.
- [90] Salar Khaleghzadegan, Sadaf Kazi, and Michael A Rosen. Unobtrusive measurement of team cognition: A review and event-based approach to measurement design. *Contemporary Research*, pages 95–113, 2020.
- [91] Christopher Best and Benjamin Rice FLTLT. Science and technology enablers of live virtual constructive training in the air domain. *Air & Space Power Journal*, 32(4):59–73, 2018.
- [92] Joseph R Keebler, Deborah DiazGranados, and Dustin C Smith. Learning team theories and measurement through the game pandemic. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 58, pages 442–446. SAGE Publications Sage CA: Los Angeles, CA, 2014.
- [93] Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. Eyeco: Effects of shared gaze on social presence in an online cooperative game. In *International Conference on Entertainment Computing*, pages 102–114. Springer, 2016.

- [94] Simo Järvelä, J Matias Kivikangas, Jari Kätsyri, and Niklas Ravaja. Physiological linkage of dyadic gaming experience. *Simulation & Gaming*, 45(1):24–40, 2014.
- [95] Michiel M Spapé, J Matias Kivikangas, Simo Järvelä, Ilkka Kosunen, Giulio Jacucci, and Niklas Ravaja. Keep your opponents close: social context affects eeg and fmg linkage in a turn-based computer game. *PLoS One*, 8(11):e78795, 2013.
- [96] Michael D. McNeese, Priya Bains, Isaac Brewer, Cliff Brown, Erik S. Connors, Jr. Tyrone Jefferson, Jr. Rashaad E.T. Jones, and Jr. Ivanna Terrell. The neocities simulation: Understanding the design and experimental methodology used to develop a team emergency management simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(3):591–594, 2005.
- [97] José M Morales, Juan F Ruiz-Rabelo, Carolina Diaz-Piedra, and Leandro L Di Stasi. Detecting mental workload in surgical teams using a wearable single-channel electroencephalographic device. *Journal of surgical education*, 76(4):1107–1115, 2019.
- [98] Gregory J Funke, April M Bennett, W Todd Nelson, and Scott M Galster. Collaboration technologies and the supervisory control of ucavs in tactical c2: Effects on performance and workload. In *2007 International Symposium on Aviation Psychology*, page 216, 2007.
- [99] James C Walliser, Patrick R Mead, and Tyler H Shaw. The perception of teamwork with an autonomous agent enhances affect and performance outcomes. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 231–235. SAGE Publications Sage CA: Los Angeles, CA, 2017.
- [100] William Bezdek, Joel Maleport, and Robert Olshan. Live, virtual & constructive simulation for real time rapid prototyping, experimentation and testing using network centric operations. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*, page 7090, 2008.
- [101] Douglas D Hodson, Bruce L Esken, Alex J Gutman, and Raymond R Hill. Quantifying radar measurement errors in a live–virtual–constructive environment to determine system viability: a case study. *The Journal of Defense Modeling and Simulation*, 11(2):115–124, 2014.
- [102] JE Hannay, OM Mevassvik, A Skjeltorp, and K Brathen. Live, virtual, constructive (lvc) simulation for land operations training: Concept development & experimentation (cd&e). In *Proc. NATO Modelling and Simulation Group Symp. on Integrating Modelling & Simulation in the Defence Acquisition Lifecycle and Military Training Curriculum (STO-MP-MSG-126)*, 2014.
- [103] Patricia Sariñana-González, Ángel Romero-Martínez, and Luis Moya-Albiol. Cooperation between strangers in face-to-face dyads produces more cardiovascular activation than competition or working alone. *Journal of Psychophysiology*, 2018.
- [104] Patricia Sariñana-González, Ángel Romero-Martínez, and Luis Moya-Albiol. Cooperation induces an increase in emotional response, as measured by electrodermal activity and mood. *Current Psychology*, 36(2):366–375, 2017.
- [105] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2013.
- [106] Thanh Nguyen, Chee Peng Lim, Ngoc Duy Nguyen, Lee Gordon-Brown, and Saeid Nahavandi. A review of situation awareness assessment approaches in aviation environments. *IEEE Systems Journal*, 13(3):3590–3603, 2019.
- [107] Koen Van De Merwe, Henk Van Dijk, and Rolf Zon. Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology*, 22(1):78–95, 2012.
- [108] Mary M Rodgers, Vinay M Pai, and Richard S Conroy. Recent advances in wearable sensors for health monitoring. *IEEE Sensors Journal*, 15(6):3119–3126, 2014.
- [109] Nasim Hajari, Irene Cheng, Bin Zheng, and Anup Basu. Determining team cognition from delay analysis using cross recurrence plot. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3482–3485. IEEE, 2016.
- [110] Jessica L Wildman, Eduardo Salas, and Charles PR Scott. cognition in teams: A cross-domain review. *Human factors*, 56(5):911–941, 2014.
- [111] Sean Andrist, Andrew R Ruis, and David Williamson Shaffer. A network analytic approach to gaze coordination during a collaborative task. *Computers in Human Behavior*, 89:339–348, 2018.
- [112] SJ Guastello. Group dynamics: Adaptability, coordination, and leadership emergence. *Chaos and complexity in psychology: Theory of nonlinear dynamical systems*, pages 402–433, 2009.