Human Factors in Virtual Environments and Game Design, Vol. 137, 2024, 50–59
https://doi.org/10.54941/ahfe1004986

Using Multi-Modal Physiological Markers and Latent States to Understand Team Performance and Collaboration

Ashley Haya Rabin[1], Catherine Neubauer[2], Kevin King[1], and Stephen M. Gordon[1]

[1]DCS Corporation, Arlington, VA 22310, USA
[2]US DEVCOM Army Research Laboratory APG, MD 21005, USA

ABSTRACT

Squads of the future battlefield will include a mixture of technically savvy humans and artificially intelligent teammates. Contextually aware AI teammates will be essential for war fighter overmatch. To understand how multimodal physiology can impact mixed team performance, we looked at how physiological team properties emerge in a naturalistic and collaborative environment. Here, we examined internal states and team outcomes based on these states within the context of a complex bomb defusal task in a simulated and naturalistic environment. This overarching research integrates eye gaze behavior, neural activity, speech, heart rate variability, and facial expressions to unravel the intricate relationship between individual and team performance. Here we focus on the facial expression data. Using a novel testbed, we aimed to uncover how these physiological processes evolve and interact with human interactions to influence team dynamics and task performance. Compared to traditional highly controlled lab tasks, this novel testbed enables peripheral measurement of multimodal physiology during naturalistic team formation and collaboration. We report differences between an individual task and teaming task in global facial expressivity results and correlations between facial expression synchrony scores and team task performance.

Keywords: Human autonomy teaming, Synchrony, Facial expression, Ecological validity, Team dynamics, Team performance

INTRODUCTION

In the future technologically advanced battlefield, successful teams must bring together diverse expertise and specialized skills. With the complexity of modern warfare involving AI, robotics, and cyber operations, interdisciplinary collaboration within teams will be critical to ensure an approach that can be fielded with little to no training (Lakhmani et al., 2022). Coordinated actions, operational flexibility, and the expertise to address ethical considerations make teams an essential organizational structure for optimizing the capabilities of advanced technologies on the battlefield. As we recognize the pivotal role of teams in the future battlefield, it becomes important to examine factors that enable positive team dynamics and ultimately enhance team performance. Beyond the diverse skills and collaborative efforts of individuals, the intricacies of teammates' physiology within teams may be able to reveal the state of the team and potentially predict the teams' performance related outcomes (Neubauer et al., 2020b).

Physiology has been used in previous research efforts to help understand a myriad of phenomena and their manifestations including stress, attention, emotion, and other cognitive states (Neubauer et al., 2020b). Some work has examined the extent to which these physiological and behavioral signals can be collapsed across teams (Berg et al., 2021; Gordon, King & Rabin, 2023; King Gordon & Rabin, 2023). More specifically, the idea that team's physiological signals can become synchronized over time may offer insight into, and provide a measure of cohesion, trust, or overall closeness. Further, facial synchrony, the coordinated mirroring of facial expressions among team members, is an example of one type of behavioral signal that holds promise as a key factor influencing team outcomes. Understanding how facial expressions contribute to the overall dynamics of teams becomes essential in optimizing their effectiveness. Here, we explore whether facial synchrony within teams exists and to what extent this may impact individual and team level performance. Drawing from the rich literature on emotional contagion and interpersonal coordination (Hatfield et al., 1994), our analyses aim to reveal the subtle, yet impactful facial expressions that may contribute to synchrony within dyads. The first layer of our investigation involves quantifying the frequency and intensity of shared emotional expressions between dyadic team members. By employing facial analysis techniques, we seek to identify synchrony characterized by simultaneous and mirroring facial expressions (Kendon, 1970).

Facial Action Units (FAUs) correspond to specific facial muscle movements, the measurement of which contribute to the analysis and computation of automated facial expression calculations. Facial Action Units were introduced as a comprehensive coding system to describe the various muscle movements involved in *typical* facial expressions (Ekman & Friesen, 1969). For example, a smile involves the activation of certain facial muscles, which can then be identified and coded using specific FAUs according to FACS (see Table 1). The system allows researchers to precisely analyze and describe facial expressions in a standardized way, providing a detailed understanding of the components that make up different emotions. FAUs have been widely used in research related to emotion recognition (Masur, Costa, Figueredo, & Teichrieb, 2023), psychology (Dawel, Miller, Horsburgh, & Ford, 2022; Neubauer et al., 2020), and human-computer interaction (Dornaika & Raducanu, 2009; Gomez, Morales, Fierrez, & Orozco-Arroyave, 2023), contributing to a more nuanced understanding of the relationship between facial expressions and latent emotional states.

Table 1. Facial expression calculation from single AUs (Ekman and Friesen, 1978).

Emotion Classification	Action Units
Anger	4+5+7+23
Contempt	R12A+R14A
Disgust	9+15+16
Fear	1+2+4+5+7+20+26
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B + 26

PROCEDURE

Face and behavioral data were collected from a total of 60 individuals (42F, 26M) with an average age of 31 years (ranging from 18-64), resulting in a total of 34 dyads. Data from 4 dyads were excluded due to missing data or data quality issues in one or more participants, resulting in a total of 30 final dyads, (36F, 24M), which were utilized in the analysis. Recruitment took place through third-party online platforms in and around the greater Los Angeles area.

Voluntary consent was obtained from all participants in accordance with Title 32, Part 219 of the CFR and Army Regulation 70-25. The Institutional Review Board of the United States Army Research Laboratory approved all human subjects testing (protocol number 21-110). Eligibility criteria included participants being 18 years or older, having normal hearing, normal (or corrected to normal) vision, normal color vision, and fluency in English. Exclusion criteria encompassed a history of brain trauma, heart problems, and pregnancy. These criteria were evaluated using a web-based prescreen questionnaire administered via Qualtrics and confirmed in lab.

Data collection occurred in pairs. Each participant in the dyad interacted with a personal experimenter and were placed in separate lab spaces away from their teammate, thus, teammates were not in the same room with one another. Once in the lab, participants underwent in-person screening, including an Ishihara color vision test and Snellin Chart (20/40). Those who did not pass this screening were excluded and compensated respectively. Participants completed a demographic battery as well as various individual and team state questionnaires. Physiological data, including EEG, heart rate, pupil size, voice, and face recordings were collected; however, for the purposes of this paper only the facial expression data and team-related performance and questionnaire outcomes are presented. Two synchronized data recording setups were used, facilitated by the Lab Streaming Layer (LSL) software (Kothe, 2014).

The primary task in the current study was the Timed Improvised Explosive Device (TIED) task presented on a 24in Dell Monitor, which was run using Unreal Engine 4. In this task, participants were required to locate IEDs scattered throughout a factory setting (see Figure 1). Upon identifying an IED, the player(s) were instructed to defuse it using the appropriate key code provided by the experimenters. Each IED was color-coded to correspond to the specific key code necessary for disarming the device. Timers on each IED would begin to countdown (even distribution of 15–30 seconds) when a participant entered a specific radius around the device. To disarm the device successfully, participants had to press and hold the correct key for 5 seconds while standing near it. Incorrect key presses or failure to disarm before the countdown finished resulted in the IED exploding. Two versions of the TIED task were employed: an individual version (i.e., TIED-I) and a team version (i.e., TIED-T). Participants began with a practice version of TIED-I involving 4 IEDs to familiarize themselves with the game dynamics and the keyboard controls.

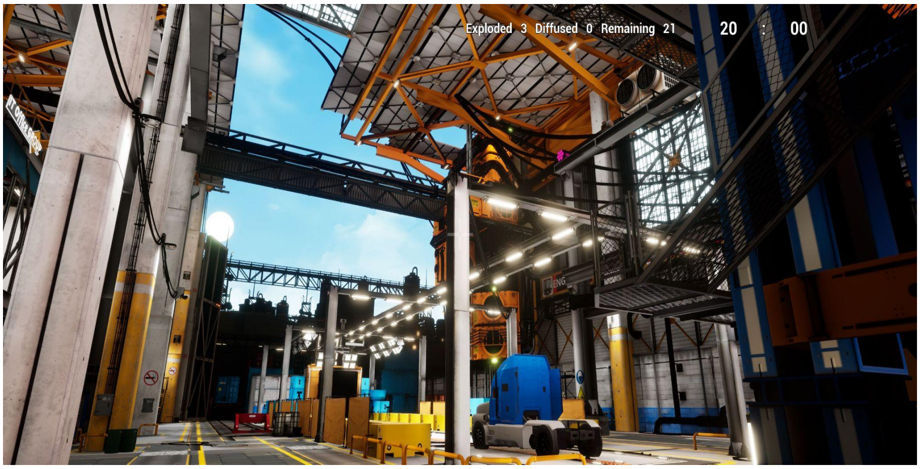

Figure 1: Factory environment used for TIED-T and TIED-I with overlay screen.

In both TIED-I and TIED-T, participants navigated the same factory environment and sought out IEDs for defusal. An overlay screen on the display provided information on game time remaining, the number of IEDs remaining, total exploded, and total disarmed. Key distinctions between TIED-I and TIED-T included the quantity and placement of IEDs, IED key codes, game duration, and the number of participants required for defusal. Specifically, TIED-I featured 16 IEDs in the environment, participants possessed key codes for all 16 IEDs, and the game lasted for 15 minutes. In contrast, TIED-T included 24 IEDs, extended gameplay to 20 minutes, and each of the two participants had access to the key codes for 12 IEDs. In TIED-T, successful defusal necessitated coordination between participants to identify who possessed the appropriate key code for a given IED, and both participants had to press the same key code simultaneously for 5 seconds.

After the completion of TIED-T, participants completed the Group Cohesion Questionnaire (GCQ) and Short Stress State Questionnaire (SSSQ). The GCQ (Carless & DePaola, 2000; Widemeyer, Brawley, and Carron) consists of 10 items measuring three factors- task cohesion, social cohesion, and group attraction. Thid questionnaires uses a 9-point Likert-type scale ranging from strongly disagree (1), to strongly agree (9). The SSSQ (Helton, 2004) is a shortened measure of stress state based on the Dundee Stress State Questionnaire (DSSQ; Matthews et al., 2002). The short measure includes factors of task engagement, distress, and worry. This is a 14-item questionnaire rated along a 5 point scale (not at all-1, A little bit-2, Somewhat-3, Very much-4, Extremely-5).

ANALYSIS PLAN

In this manuscript, our focus is directed specifically towards a subset of data, including facial expression data, while the comprehensive analysis plan and results of questionnaires and physiological signals fall beyond the scope of the current paper. The detailed examination of these additional elements, including physiological signals, will be presented separately. The following section will provide results from global facial expressivity analysis, followed by measures of global synchronicity for individuals and teams as they proceeded through the task, as well as correlations between synchronicity, performance, and subjective state measures.

Analysis of Facial Expressivity

The participant's face was continuously recorded throughout the task via a webcam mounted to the simulation screen. Measures relating to emotional expression were automatically extracted through the OpenFace freeware (Baltrušaitis et al., 2018). More specifically, OpenFace yields frame-by-frame evidence of facial action unit (AU) evidence, which corresponds to specific muscle movements of the face yielding values on a scale of 0 – 5. Facial expressions relating to both positive and negative affect (i.e., emotions such as happiness, sadness, surprise, fear, anger, and contempt) were calculated on a frame-by-frame basis separately for each TIED-T and TIED-I, using computations of single AU evidence following the Facial Action Coding System described in previous sections (Ekman and Friesen 1978). However, it should be noted that these calculations are based solely on action unit movement in the face, and while evidence of AU changes have been linked to universal emotions in humans, the analysis of expressivity warrants further investigation and should be combined with other latent state measures for ground truth and cross-referencing of observable human emotion.

Initially, we examine variations in facial expressivity between TIED-I, executed individually, and TIED-T performed collaboratively within a team. Subsequently, we delve into the correlation between synchrony and team performance. Finally, an exploration of covariates is conducted, considering both individual-level states and team-level states.

RESULTS

First facial expressivity was compared between the two TIED tasks. A one-way between subjects ANOVA was conducted to compare the averaged facial expression scores (i.e., happiness, sadness, surprise, fear, anger, disgust, and contempt) when participants worked individually or within their team. Overall, the descriptive statistics on the averaged expression data, revealed that evidence of facial expressivity was somewhat higher across all seven expressions when participants worked in a team, compared to when they worked alone.

Table 2. Differences in mean evidence of facial expressivity in individual vs. team scenarios. An * indicates significant differences between individual vs. team at the 0.01 level, ** indicates significance at the 0.001 level.

Facial Expression	Individual vs. Team	Mean	Std. Deviation
Happiness	Individual	0.29**	0.25
	Team	0.56**	0.39
Sadness	Individual	0.23	0.13
	Team	0.24	0.12
Surprise	Individual	0.18*	0.05
	Team	0.21*	0.06
Fear	Individual	0.22	0.11
	Team	0.24	0.09
Anger	Individual	0.22	0.17
	Team	0.23	0.15
Disgust	Individual	0.14	0.06
	Team	0.16	0.06
Contempt	Individual	0.65**	0.09
	Team	0.95**	0.34

More specifically, results revealed several significant differences for the following expressions: happiness $F(1, 129) = 22.20$, $p < 0.001$, surprise $F(1, 129) = 8.77$, $p < 0.01$, and contempt $F(1, 129) = 14.81$, $p < 0.001$, again indicating that expressivity was higher for these states when participants worked in a team vs individually.

Next, synchrony scores were computed for each team by borrowing a linguistic synchrony formula (See below; Baker et al., 2021). Equation 1 uses the seven facial expression scores from each player (A & B) and determines how similar they are to each other. Synchrony scores are on a scale from 0 to 1 with values closer to 1 reflecting higher synchrony.

$$\frac{\sum_i^7 \left(1 - \frac{|Data\_Set\_A_i - Data\_Set\_B_i|}{\sum_j^7 (|Data\_Set\_A_j + Data\_Set\_B_j|)}\right)}{7} \tag{1}$$

Correlations were then computed between synchrony score and performance metrics collected during TIED-T (See Table). We found a significant positive correlation between facial synchrony and total number of defused IEDs ($r = 0.33$, $p < 0.01$), as well as total IEDs found ($r = 0.31$, $p < 0.01$). Indicating that as synchrony scores between teammates increased, performance with IEDs found in the TIED-T increased. Interestingly, synchrony scores were negatively correlated with a sub-factor measure of *Engagement* ($r = -0.27$, $p<0.05$). A significant negative correlation with age was also found such that as age increased, synchrony decreased (-0.26, $p < 0.05$). Lastly, we saw a positive correlation between synchrony and a subfactor of the Group Cohesion Questionnaire, *Group Attraction* ($r = 0.26$, $p < 0.05$).

Table 3. Summarized results with facial synchrony score and p score.

Variables	r	p
Age x Synchrony Score	-0.26	0.045
IEDs defused x Synchrony Score	0.33	0.003
IEDs found x Synchrony Score	0.31	0.015
Engagement x Synchrony Score	-0.27	0.035
Group Attraction x Synchrony Score	0.26	0.045

CONCLUSION

Individual vs Team Facial Expressivity

Our investigation into facial expressivity during individual (TIED-I) and team (TIED-T) tasks revealed notable differences. We compared the averaged facial expression scores of happiness, sadness, surprise, fear, anger, disgust, and contempt. Our results indicated a generally higher level of facial expressivity across all expressions when participants worked in TIED-T compared to TIED-I. Specifically, significant differences were observed for happiness, surprise, and contempt, signifying increased expressivity in these states during team tasks.

These analyses did not enable us to extract why some expressions were significantly higher compared to other expressions in TIED-T. Our task took place in a VR rendered factory environment in both the individual and team tasks and contained identical game dynamics. It could be that the sole act of teaming with strangers increases feelings of happiness, surprise, and contempt. There is also the potential for individuals to increase displayed indicators of happiness, surprise, and contempt to communicate or bond with teammates. Though in our experiment, participants' facial expressions were not displayed and/or visible to their teammates. Future designs could compare differences in facial expressivity when teams can see their teammates faces during the task, versus not to further elucidate the underpinnings of these expressions in teaming contexts.

Facial Synchrony and Team Outcomes

In our exploration of team dynamics, we focused on team synchrony during TIED-T. The linguistic synchrony formula enabled the computation of a synchrony score using facial expression data from each team member. Correlation analyses were conducted between synchrony scores and performance metrics during TIED-T. The results revealed significant positive correlations between facial synchrony and the total number of found and defused IEDs (See below). These findings suggest that heightened synchrony is associated with improved performance in identifying and defusing IEDs during team tasks. This is a particularly compelling finding given that both performance metrics were dependent on the team as a whole. In TIED-T each member had half of the codes and to defuse the IED teams were required to collaborate on multiple fronts. Teams first needed to decipher what the correct shade of the IED was, who had the code for that shade, and press the correct code in the correct location at the same time for five seconds while both standing in the correct proximity to the IED, while under time stress of an IED countdown. Taken together, these task factors illustrate the complexity of the collaboration needed to perform well at this task. More analyses are warranted to understand the degree to which synchrony can explain the variance across performance, though these results are still encouraging. Given the search-plus-action structure of the TIED-T task, the impact of physiological synchrony should be explored in other testbeds and compared across other physiological signals.

Significant negative correlations emerged between synchrony scores and the *Engagement* sub-factor of the Short Stress State Questionnaire ($r = -0.27$, $p < 0.05$). This sub-factor, encompassing motivation, concentration, and energy directed toward the task (Mathews et al., 2002), suggests that participants' focus on their teammate may surpass their focus on the task itself. Given the collaborative nature of the task, participants may prioritize interpersonal dynamics over the specific task of locating and defusing IEDs. Exploring whether an emphasis on team cohesion, rather than task performance, contributes to enhanced overall task performance warrants further investigation. Further, the relationship between *Engagement* and synchrony is particularly interesting given the relationship between synchrony and the Group Environment Questionnaire subfactor *Group Attraction* ($r = 0.26$, $p < 0.05$). Group Attraction captures a given teammates' impressions of social interactions and personal acceptance within the team (Whitton & Fletcher, 2014). The positive relationship between Group Attraction and synchrony indicates that the more individuals perceive positive social interactions and personal acceptance from their teammate, the higher their synchrony. This finding underscores the importance of social cohesion and positive team dynamics in influencing non-verbal coordination, as reflected in synchrony, among team members during the TIED-T task. It suggests that fostering a positive team environment, where individuals feel socially connected and accepted, may contribute to enhanced synchrony and may improve overall team performance.

Lastly, we uncovered a negative relationship between age and synchrony, in that as age increased, synchrony decreased ($r = -0.26$, $p < 0.05$). Older individuals may have less gaming experience compared to their teammates and may have spent more energy focused on understanding the game dynamics and task requirements, rather than team cohesion. Future work should examine how game use, individual age, and teammate age gaps may contribute to synchrony to further unpack this relationship.

Limitations and Future Directions

Coupled together, these overarching findings underscore the intricate relationship between facial synchrony and key performance metrics in team scenarios. The negative correlations with stress-related engagement factors and age suggest potential implications for team dynamics and individual engagement. However, the positive correlation with group attraction prompts further exploration into the nuanced dynamics influencing team cohesion.

It is important to note that our analysis focused solely on synchrony, and while informative, future research should delve into the specific mechanisms that drive these correlations. One promising avenue could be to explore whether analysis of other physiological signals such as pupilometry or electroencephalography mirror the current results. Additionally, considering the limitations of our experiment, future designs could explore the impact of visual cues, such as displaying facial expressions, on team dynamics during tasks. This could provide valuable insights into the communicative aspects of facial expressions within a team context.

We described the role of team dynamics in future military scenarios, emphasizing the potential facilitation of team deployment through physiological considerations. The literature review explored facial expression in anticipation of team outcomes. The study detailed a comprehensive collection of behavioral, physiological, and self-report variables to delineate the experimental scope. Our focus on synchrony emerged as a central theme for understanding and predicting team performance. Our findings indicated a noticeable increase in facial expressivity during collaborative tasks. Importantly, we substantiated the association between team synchrony and improved performance outcomes.

This research contributes to the field of team dynamics and performance prediction, offering insights into the interplay of physiological, behavioral, and self-report factors in collaborative and naturalistic testbed.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Bianca Dalangin for her assistance in data collection. Special thanks to Robert Smith, Adam Wachsman, and Hailey Johnson for their contributions to the development of the TIED tasks.

REFERENCES

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2018). OpenFace: An open-source facial behavior analysis toolkit. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 1–10.

Baker, A. L., Fitzhugh, S. M., Forster, D. E., & Schaefer, K. E. (2021, September). Communication Metrics for Human-Autonomy Teaming: Lessons Learned from us Army Gunnery Field Experiments. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 65, No. 1, pp. 1157–1161). Sage CA: Los Angeles, CA: SAGE Publications.

Berg, S., Neubauer, C., Robison, C., Kroninger, C., Schaefer, K. E., & Krausman, A. (2021, June). Exploring Resilience and Cohesion in Human-Autonomy Teams: Models and Measurement. In International Conference on Applied Human Factors and Ergonomics (pp. 121–127). Cham: Springer International Publishing.

Carless, S. A., & DePaola, C. (2000). The Group Environment Questionnaire: A test of the revised version. Small Group Research, 31(6), 719–736.

Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000–2020. Behavior Research Methods, 54(4), 1889–1901.

Dornaika, F., & Raducanu, B. (2009). Facial expression recognition for HCI applications. In Encyclopedia of Artificial Intelligence (pp. 625–631). IGI Global.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1(1), 49–98.

Ekman, P. (1982). Methods for measuring facial action. Handbook of methods in nonverbal behavior research, 45–90.

Files, B. T., Oiknine, A. H., Thomas, J., Khooshabeh, P., Sinatra, A. M., & Pollard, K. A. (2020). Same task, different place: Developing novel simulation environments with equivalent task difficulties. In Advances in Human Factors and Simulation: Proceedings of the AHFE 2019 International Conference on Human Factors and Simulation, July 24–28, 2019, Washington DC, USA 10 (pp. 108–119). Springer International Publishing.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K.,... & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. Emotion, 2(4), 315.

Gomez, L. F., Morales, A., Fierrez, J., & Orozco-Arroyave, J. R. (2023). Exploring facial expressions and action unit domains for Parkinson detection. Plos one, 18(2), e0281248.

Gordon, S., King, K., & Rabin, A. (2023). Latent State Synchronization in Dyadic Partners using EEG. In the 2023 IEEE Systems, Man, and Cybernetics Conference.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). Emotional contagion. Cambridge University Press.

Helton, W. S. (2004). Short Stress State Questionnaire. Behavioural Research Methods, Instruments, & Computers, 36(4), 600–604.

Kendon, A. (1970). Movement coordination in social interaction: Some examples described. Acta Psychologica, 32(2), 100–125.

King, K., Gordon, S., & Rabin, A. (2023). Granger leadership as a surrogate for task engagement in a novel dyadic search paradigm. In the 2023 IEEE Systems, Man, and Cybernetics Conference.

Kothe, C. (2014). Lab streaming layer (LSL).

Lakhmani, S. G., Neubauer, C., Krausman, A., Fitzhugh, S. M., Berg, S. K., Wright, J. L., & Schaefer, K. E. (2022). Cohesion in human–autonomy teams: An approach for future research.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K.,... & Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion, 2(4), 315. Oretical Issues in Ergonomics Science, 23(6), 687–724.

Masur, P., Costa, W., Figueredo, L. S., & Teichrieb, V. (2023). Leveraging Previous Facial Action Units Knowledge for Emotion Recognition on Faces. arXiv preprint arXiv:2311.11980.

Mayo, O., & Gordon, I. (2020). In and out of synchrony—Behavioral and physiological dynamics of dyadic interpersonal coordination. Psychophysiology, 57(6), e13574.

Neubauer, C., Schaefer, K. E., Oiknine, A. H., Thurman, S., Files, B., Gordon, S., & Gremillion, G. (2020b). Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human Autonomy Teaming. CCDC Army Research Laboratory Aberdeen Proving Ground United States.

Neubauer, C., Gremillion, G., Perelman, B. S., La Fleur, C., Metcalfe, J. S., & Schaefer, K. E. (2020, January). Analysis of facial expressions explain affective state and trust-based decisions during interaction with autonomy. In International Conference on Intelligent Human Systems Integration (pp. 999–1006). Cham: Springer International Publishing.

Whitton, S. M., & Fletcher, R. B. (2014). The Group Environment Questionnaire: A multilevel confirmatory factor analysis. Small Group Research, 45(1), 68–88.