ARL-TR-9070 ● Sep 2020

Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human–Autonomy Teaming

by Catherine Neubauer, Kristin E Schaefer, Ashley H Oiknine, Steven Thurman, Benjamin Files, Stephen Gordon, J Cortney Bradford, Derek Spangler, and Gregory Gremillion

NOTICES

Disclaimers

ARL-TR-9070 ● SEP 2020

Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human–Autonomy Teaming

Catherine Neubauer, Kristin E Schaefer, Steven Thurman, Benjamin Files, J Cortney Bradford, Derek Spangler, and Gregory Gremillion
Human Research and Engineering Directorate, CCDC Army Research Laboratory

Ashley H Oiknine and Stephen Gordon
DCS Corporation

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED (From - To)**
September 2020	Technical Report	November 2018–September 2020

4. TITLE AND SUBTITLE Multimodal Physiological and Behavioral Measures to Estimate Human States and Decisions for Improved Human–Autonomy Teaming	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Catherine Neubauer, Kristin E Schaefer, Ashley H Oiknine, Steven Thurman, Benjamin Files, Stephen Gordon, J Cortney Bradford, Derek Spangler, and Gregory Gremillion	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) CCDC Army Research Laboratory ATTN: FCDD-RLH-FA Aberdeen Proving Ground, MD 21005	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-9070
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ORCID ID(s): Catherine Neubauer, 0000-0002-6686-3576; Kristin E Schaefer, 0000-0002-1342-3446; Ashley Oiknine, 0000-0001-9092-7958; Steven Thurman 0000-0003-3962-8447; Benjamin Files 0000-0002-1141-7886; J Cortney Bradford 0000-0002-5926-9229; Derek Spangler 0000-0001-7172-9083; Gregory Gremillion, 0000-0002-0205-688X

14. ABSTRACT
For effective human–autonomy teaming to occur, it is essential to manage the interactions between humans and autonomous agents. This leads to optimized performance, ensures resilient and stable team member states, and supports the capability to appropriately build, calibrate, and maintain trust. Toward this aim, it is critical to predict human teammate decisions, and the underlying mental states that drive those decisions. By predicting these decisions, we will be able to design new intervention strategies and technologies, such as display designs, agent feedback, or adaptive behavior, to improve teaming and mitigate possible negative interactions such as performance degradations and miscalibrated trust. In this report, we motivate the importance of estimating the psychological states that impact these decisions and summarize the known relationships they have with physiological and behavioral measures that can be captured in real time with noninvasive or wearable technologies. This provides a foundation to employ a priori constraints on models that utilize multiple physiological and behavioral signals to infer mental or psychological states (stress, fatigue, workload, trust, etc.), to improve prediction of human decisions when interacting with autonomous agents in military-relevant environments.

15. SUBJECT TERMS
multimodal psychophysiology, human–autonomy teaming, affective computing, trust-based decision-making, state estimation, human computer interaction, Human–Autonomy Teaming Essential Research Program, HAT ERP, Next Generation Combat Vehicle, NGCV

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Catherine Neubauer
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified	UU	82	(310) 448-5354

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Summary

The purpose of this report is to provide a resource that characterizes and summarizes relationships between observable measures of human physiology and behavior with unobservable mental states, to support multimodal inference informed by those relationships. By taking this approach, we argue it is possible to generate mutually informed estimates of those unobservable states and improve predictions of interactions between humans and autonomous counterparts. This will enable the capability to more accurately predict joint decision-making, which is needed to identify and initiate appropriate state-based interventions for effective and appropriate human–autonomy teaming. This directly supports the US Army Combat Capabilities Development Command Army Research Laboratory's Human–Autonomy Teaming Essential Research Program and the Next-Generation Combat Vehicle mission prioritization by supporting new techniques for predicting human states and behaviors related to decision-making.

Within this context, various approaches, including machine learning and classical filtering and estimation techniques, are being explored as a means to estimate and predict human mental states that can inform effective human–autonomy team interactions and performance. This view is motivated by the principle that in a human–autonomy team, the human can, and often should, be thought of as a sensor that can provide critical information to the joint decision-making process, or as an element whose time-varying state should be considered to optimize team performance and resilience. Understanding fluctuations in human state are important because they provide continuous insight into the processes that drive their interactions with teammates. This also builds on previous research utilizing wearable sensor technologies to infer underlying human states, such as stress, fatigue, workload, trust, and others, which can directly impact how human team members make decisions as they interact with autonomous technologies.

Here, we look at how employing a priori knowledge of these relationships can constrain models and be used to empirically fit optimal, interpretable, and plausible predictions to physiological and behavioral data when multiple sensor modalities are available. This allows us to use multiple human-derived signals to better model their relationships to a common set of mental states, rather than approaches that aim to estimate those states from a single modality in isolation. The ability to leverage information from multiple observable sensor sources allows us to more robustly infer unobservable latent states, which are informed and constrained by domain knowledge. This approach can provide more reliable and interpretable estimates of states that impact the resultant interactions with autonomous teammates, which we aim to improve. Therefore, this report collects and synthesizes the informative

relationships between various human-centric sensor modalities and the mental states that impact teaming. The focus on characterizing these relationships for a set of wearable, noninvasive sensing mechanisms supports Army goals of fieldable methods for real-time Soldier state estimation to dynamically manage interactions in human–autonomy teams.

1. Introduction

Current Army missions are characterized by an increased reliance on automation to extend human capabilities, creating a growing need to understand the decisions that impact human–autonomy interactions. A gap exists in current technologies for explicitly modeling and predicting the psychological mechanisms that drive this interaction, making optimal integration of human–autonomy teams difficult. Human assessments that combine wearable technologies with advanced analytics can provide a deeper understanding of an individual's psychophysiological responses, including predictions of their resultant actions (Gremillion et al. 2018; Marathe et al. 2020). These responses can be used to characterize their interaction with autonomous technology and provide quantifiable data that can be used to assess teaming processes. This is important because human–autonomy teams are steadily evolving from user–tool relationships to that of teammates. Increasingly social, fluid, and naturalistic interactions are needed, whereby coordinated actions and joint decisions made by humans and autonomous counterparts (software systems or embodied robots) require the maintenance of shared mental models and real-time shared situation awareness of both the task and other team members (Phillips et al. 2011).

As such, advances in wearable devices have enabled the tracking of a wide range of factors including activity, sleep patterns, and various physiological response (Bonato 2010). These characteristic quantities can provide critical insight into factors that influence the human–autonomy joint decision-making process, such as fatigue, stress, workload, vigilance, and trust. While the process of effectively integrating a human as a sensor to better predict joint human–autonomy decision-making is a relatively new field of study, this report argues for the use of psychophysiological signals to enabling multimodal sensor fusion, improved algorithms for assessing human state, and a cue mechanism for adapting and managing team interactions. This report motivates the approach of multimodal human-centric sensing to improve human–autonomy interactions (Section 1), outlines the metrics used in human state estimation (Section 2), describes specific physiological and behavioral sensing modalities that can be leveraged to that end (Section 3), and discusses the specific latent states deemed especially relevant to decision-making within this context (Section 4).

1.1 Using Physiological Signals to Manage Human–Autonomy Team Interactions

Advances in autonomous technologies have been designed to improve performance of the human operator by creating a less stressful, more effective working environment, which include functionalities of autonomous technologies that are adaptive to the human user (Desmond and Hancock 2001). Thus, the goal of teaming with autonomous systems is to relieve the human of tasks that are tedious; taxing on their mental, cognitive, or physical state; or not suited to their relative strengths. As such, many researchers have suggested that task modes, levels of autonomy, or control authority be automatically varied based on changes in the real-time state of the human (Parasuraman et al. 1992; Scerbo 1996; Metcalfe et al. 2017; Gremillion et al. 2018). For example, physiological signals that reflect changes in autonomic function (and presumably workload, stress, or arousal) could serve as a trigger for adaptive task allocation, to offload tasks to available teammates, when these signals are above or below desirable performance and decision-making thresholds (Morrison and Gluckman 1994; Byrne and Parasuraman 1996; Kramer et al. 1996). Here, physiological states may signify a precursor to maladaptive decision-making or performance impairments, which could be used to provide a notification, warning, or alert; a change in teammate communication or behavior; or an adjustment to the level of autonomy accordingly.

Although the advancements of autonomous systems have been widely documented, there are particular concerns to be considered when forming human–autonomy teams, especially in Army-relevant domains. First, autonomous systems may induce complacency or poorly calibrated expectations about the capabilities of those systems and increase the likelihood of the human to over- or under-rely on them. Second, these technological advancements may therefore produce minimally beneficial or even detrimental effects if the human team member does not have an appropriate level of trust in the system. Third, autonomous systems may interact with individual differences of their human teammates to produce highly variable changes in physiological responses and behaviors relating to the use of those system. Therefore, to effectively manage interactions in a human–autonomy team, it is crucial to model the real-time dynamics of the human team member's mental states associated with decision-making and understand their impact on those interactions.

1.2 Assessing Human State

To accomplish the aforementioned task, it is essential to employ metrics that continually gauge critical human team member states to inform the necessary configurations, actions, or interventions to encourage appropriate teaming. Here, state estimation may be enhanced by developing algorithms that integrate information from multiple sensors, which can include both noninvasive technologies relating to computer vision approaches and wearable sensors capable of measuring physiological responses (Sahayadhas et al. 2013). Given that different state indices are not always highly correlated, this appears to be a well-founded strategy; however, before discussing the importance of state mapping, it is important to define what is meant by state.

In terms of classical state space modeling, system states refer to the set of variables that sufficiently characterize the dynamics of a system with respect to its input–output behavior (Dukkipati 2006). For example, an embodied robotic system's physical state might often be sufficiently represented by its position and orientation, and perhaps some number of temporal derivatives of these variables. Similarly, system inputs, such as an exogenous force, describe external quantities that alter the propagation of that state over time (i.e., the system dynamics, which is represented by changes in the state, for example, the robot's position, orientation, and velocities). The states may then manifest observable system outputs via transduced measurements (e.g., the robot's gyroscope or accelerometer signals), which provide observable indicators of the system state. It should be noted that the state is not necessarily fully observed in the outputs but does necessarily capture, in observable and latent variables, the quantities needed to propagate the system forward in time.

Similarly for a human interacting with a teammate, whether human or autonomous agent, we argue that the relationship between the external stimuli they experience and the resulting output behavioral and physiological responses are governed by psychological processes that can be modeled and expressed in terms of a discrete set of latent states that represent abstract mental, emotional, and cognitive features (e.g., fatigue, stress, workload, and trust). Thus, we purport that these states are driven by external features of the environment and task stimuli and can be deduced from real-time physiological and behavioral measures. This is in line with research on task-induced states, which suggest that factors like stress, fatigue, workload, and trust predominantly moderate how humans appraise and cope with tasks demands in performance-based settings (Matthews et al. 2012b).

1.3 State Modeling, Estimation, and Prediction

Given the complex relationship between external environmental stimuli and the physiological and behavioral response of the human, a promising approach to model this relationship is to use fully data-driven methods for generalized function approximation, such as neural networks, to predict resultant responses from a relatively large set of measurable quantities. Such approaches are powerful tools for identifying nonlinear mappings from inputs to outputs, particularly in deep neural networks that incorporate multiple layers and many internal states (i.e., nodes). Applying these techniques successfully assumes access to a sufficiently large and often well-labeled data set in order to robustly fit the parameters (i.e., weights) of the network. In some applications, particularly in highly complex and variable circumstances, such as military operations, acquiring a sufficient amount of data can be intractable. Further, these models generally do not produce statistical measures of uncertainty or interpretable results, as the source of predictions are obscured by the complex internal activation state. This characteristic can make diagnosis, validation, and verification of the outputs of these models challenging.

In cases where the limitations of machine learning models (e.g., limited explainability and large data volume requirements) are prohibitive in generating useful or predictive models, approaches that use domain knowledge to constrain the model structure and complexity to a greater degree are an attractive approach. Utilizing well-characterized relationships among latent mental, emotional, and cognitive states and measurable physiological and behavioral features provides a pathway to design more compact and interpretable models whose structure is well supported by established literature. This approach can compensate for the loss of prediction accuracy that generally results from more simplified models. By encoding these relationships in the structure of the model, we can also make plausible claims about the internal human state at each moment in time with the goal of tracking their internal psychological dynamics and improving the prediction of their future behavior as part of a human–autonomy team. Machine learning and data-driven methods can therefore leverage these known domain constraints and help mitigate these limitations. At the most basic level, the selection of learning approach and architecture (e.g., fully connected versus convolutional networks, or feedforward versus recurrent networks) can be informed by the fundamental processes that govern the relationships between input and output features.

Several recent learning approaches may prove to be apt techniques for integrating psychophysiological domain knowledge to inform modeling and prediction of human decisions and actions to a greater degree. This may be is accomplished by constraining the internal structure or input-output relationships using this domain-specific a priori information. For example, promising learning methods include constrained variational autoencoders or generative networks that have a compact, latent representation. This produces an intermediate network representation with elements that, unlike traditional autoencoders, are disentangled and can be independently varied to yield explainable and interpretable changes in the network outputs along recognizable dimensions (Chen et al. 2016; Higgins et al. 2017). What this suggests is that if the elements of this latent representation can be plausibly assigned to psychological state dimensions, possibly through semi-supervised data labeling by a human expert, these pretrained models can be used to estimate psychological state by monitoring this latent layer subject to the observed time-varying network inputs (i.e., environmental stimuli). Another nonparametric approach that shows promise in applying psychophysiological and behavioral domain knowledge is constrained variational inference (Unhelkar and Shah 2019). This technique constrains the human behavioral dynamics with known relationships between latent psychological states and physiological and behavioral measures as identified by sparse, semi-supervised labels provided by a human expert. This approach differs from neural network technique in that it utilizes a Markov model framework, allowing for Bayesian optimality in estimating these behavioral policies, and integrates sequences of observed outputs, which can provide greater explainability, temporal consistency, and data efficiency.

Conversely to these nonparametric techniques, the application of first principles domain knowledge to directly define parametric system dynamics describing the human psychological processes may provide even greater benefits in terms of data efficiency and explainability. The use of traditional dynamic systems representation and recursive inference are also promising approaches to leveraging prior domain knowledge for modeling and estimating the human's latent mental states from observable measures (i.e., physiology and behavior). Dynamic systems allow for compact and interpretable representation of the unobservable state of the human. This representation defines the relationships among input, output, and state variables in the form of mathematical functions, typically state-space differential equations. Again, these dynamical equations define the temporal processes that govern forward propagation of the human state in time. The specification of these variables and their associated functional relationships does require potentially challenging a priori design, which entails a greater modeling effort than is typical in more generalized machine learning approaches. However, it allows for directly specifying plausible relationships and constraints based on existing domain knowledge. Assuming that a set of input, output, and state variables can be defined, there does exist an extensive set of system identification tools to empirically model the functional relationships among the variables (Keesman 2011; Nelles 2013). The use of compact, parameterized functional relationships also allows for flexibility to adapt to human variability through empirical fitting of those parameters with greater data efficiency than is possible with learned models that are less constrained. An additional benefit of using state-space dynamic models, is the ability to apply recursive inference techniques, which produce statistically optimal state estimates and compute levels of uncertainty in those estimated states (Simon 2006). This characteristic also acts to continuously correct errors in the latent state estimate based on real-time measurements of several output quantities (i.e., physiological and behavioral features [Gremillion et al. 2018]).

Here, we suggest that a multimodal approach can more accurately inform those estimates by combining measurements of many such linkages to simultaneously and complementarily corroborate the human state. These characteristics also allow us to apply existing techniques for optimal model predictive control, which leverages forecasts of future system states, to more effectively mitigate or drive human behavior toward a desired outcome through feedback (Allgöwer and Zheng 2012; Camacho and Alba 2013).

Independent of the particular modeling approach used, we argue that estimation of explainable, latent psychological states by integrating measurements from several physiological and behavioral modalities (that are jointly related to those states) will produce more tractable, informative estimates to improve predictions of and interventions in human decisions and actions. To enable these approaches, it is critical that the robust multimodal relationships between latent states and physiological and behavioral measures be characterized to the extent that they are well established in previous literature. Therefore, the goal of this report is to collect and summarize this information in a form that is accessible to individuals applying these multimodal modeling and estimation techniques.

2. Measures Used for Human State Estimation

The following sections describe the various tools available for assessing human state, which include 1) subjective measures, 2) performance measures, and 3) physiological measures and behavioral measures. The relationships outlined in these sections describe the links between various observable measures and unobservable states.

2.1 Subjective Measures

Traditional measures of human states have relied heavily on subjective ratings of individual affect and cognition. Typically, responses are recorded, via self-report questionnaires, both before and after a task to gauge changes in state such as changes in task-induced stress, fatigue, or trust. Several factors make questionnaires beneficial to researchers, including the low cost and ease of distribution, though there are some critical disadvantages. First, individuals may not always be able to accurately assess their thoughts and feelings, which can result in biased characterizations of their cognitive and affective state. Individual differences in motivation or other factors (i.e., compliance) may also affect the reliability of subjective responses (Thurman et al. 2018). Additionally, reporting on a potentially sensitive subject matter may hinder honest and forthcoming responses. Perhaps most important for state estimate and decision-making modeling is the fact that questionnaires ask respondents to remember how they were thinking or feeling *after* the moment of interest has occurred, at a singular time point. This method of probing the human's internal state, often collected after some significant time delay, provides only a sparse measure of state and can yield results that are biased by subjective experience and sensitive to artifacts of memory recall. Conversely, the use of noninvasive or wearable sensors for continuous, multimodal sensor fusion and estimation of human state is less susceptible to these limitations.

2.2 Performance Measures

In addition to self-reported data, performance measures can also be used to gauge changes in human state (e.g., an emergent fatigue response while performing a sustained attention task [Warm et al. 2008a, 2012]). In this context, much research has focused on what is referred to as the *vigilance decrement*, which is a psychological phenomenon that occurs during long periods of sustained attention and is characterized by a steep then steady decline in performance over time (Mackworth 1948; Matthews and Desmond 2001; Warm et al. 2008b). Additionally, the dynamic model of stress and sustained attention distinguishes between two qualitatively different fatigue states: active and passive fatigue (Desmond and Hancock 2001). Within this model, active fatigue is the result of sustained high workload tasks, which result in a depletion of cognitive resources. Conversely, passive fatigue is a product of low-workload, monotonous task environments (e.g., operating in an automated environment for an extended period of time) and results in lower levels of task engagement and a subjective withdrawal of task-related effort. Although each form of fatigue is elicited from different task parameters and result in qualitatively different states changes, both result in declines in performance. Declines in performance can also be measured in task paradigms that include a secondary task (e.g., a psychomotor vigilance task) to induce changes in workload, stress, or fatigue, which naturally detracts from the primary attention stream. In these cases, the individual will typically continue to apply effort to a primary task even after fatigue has set in (Mascord and Heath 1992).

Other performance metrics such as reaction time to standard and emergency events can also be gauged to infer specific human states such as passive and active fatigue (Saxby et al. 2008; Neubauer et al. 2012; Körber et al. 2015). In this context, several driving studies have had participants avoid collision with a vehicle that suddenly appeared from the side of the road moments after a driver had regained manual control of an automated vehicle. They found that this type of task resulted in different behavioral outcomes that were the result of a passive fatigue induction

(Saxby et al. 2008; Neubauer et al. 2012). In cases such as these, state changes can adversely impact performance.

Although useful for inferring changes in state, performance metrics first require a decrement to occur to conclude a change in state has actually happened (Matthews and Desmond 2002). Additionally, performance may suffer before individuals are cognitively aware of and can correct these deficits. Ideally, future metrics would be designed to detect harmful states and preemptively intervene *before* they result in a performance decrement or undesirable decision-making scenario. This would require the continuous, noninvasive assessment of largely unobservable states that drive human decision-making and behavioral outputs.

2.3 Physiological Measures

Methods of inferring unobservable (i.e., latent) states can be individualized to the specific response measures that are most informative to the specific context. For example, the fatigue state can be detected psychophysiologically (Wohleber et al. 2016) through well-known metrics, such as increasing spectral power in the lower frequency bands of the electroencephalogram (EEG; Borghini et al. 2014) and the percentage of eye closure time (PERCLOS; Wierwille et al. 1994). Additionally, electrocardiogram (ECG) studies have utilized decreased heart rate (HR) as a measure of lower arousal, and hence, of fatigue (Borghini et al. 2014). Relatedly, large fluctuations in pupil diameter captured by the pupillary unrest index (PUI) have been shown to index sleepiness-related fatigue (Lüdtke et al. 1998). Thus, fusing two or more physiological modalities can provide corroborative information about the unknown psychological fatigue state in this context. Therefore, we can then leverage the findings of relevant psychophysiological literature such as this to inform a model that captures all of these effects and fuses them to more robustly estimate the states that they each aim to infer.

However, the physiological responses that may be associated with underlying latent human states are only useful to the extent that they allow researchers to index a psychological process. Therefore, before describing the various methods available for human-state detection, a few limitations to this approach should be considered. Generally, it is thought that a person is anxious or stressed because they exhibit physiological activation, are fatigued or inattentive due to diminished activation, surprised because they show a startle response, and so on. However, it cannot be definitively concluded that a psychological state has been detected simply because a physiological response, previously found to vary as a function of psychological processing or a latent state, has been observed. Therefore, a general framework for viewing the relationship between both psychological and physiological events is to consider these two groups of events as representing independent sets or domains (e.g., a collection of elements who together are considered a whole), where all elements within the psychological set are assumed to have some sort of physiological representation (Cacioppo and Tassinary 1990; Cacioppo et al. 2000). More specifically, Cacioppo et al. (2007, p. 8–9) have outlined several possible domain or set relationships between the psychological and physiological sets, which are as follows:

1) "A one-to-one relation, such that an element in the psychological set is associated with one and only one element in the physiological set, and vice versa.
2) A one-to-many relation, meaning that an element in the psychological domain is associated with a subset of elements in the physiological domain.
3) A many-to-one relation, meaning that two or more psychological elements are associated with the same physiological element.
4) A many-to-many relation, meaning two or more psychological elements are associated with the same (or an overlapping) subset of elements in the physiological domain.
5) A null relation, meaning there is no association between an element in the psychological domain and that in the physiological domain."

Cacioppo and Tassinary (1990) suggest that only the first and third of these possible relationships allow for a robust specification for a psychophysiological relationship to be determined. Therefore, within the field of psychophysiology, one emphasis is on integrating multimodal data streams in order to illuminate latent psychological functions and mechanisms, rather than strict physiological structures per se. However, if a multimodal set of physiological measurements can be collected that have overlapping connections to a common set of psychological states, then the second and fourth of these relationships are able to provide valuable formation for inference of those states. The following sections discuss the utility of employing multimodal sensors to assess the relationship between specific physiological measures and their associated human states to identify the potential relationship these have on decision-making.

3. Leveraging Multimodal Measures of Psychophysiology and Behavior: The Whole is More Than the Sum of Its Parts

Individualized sensors, whether wearable or noninvasive, can reveal specific aspects of human cognition, emotion, and behavior. Such sensors are capable of capturing, to name a few, eye behaviors, EEG signals in the brain, autonomic activity relating to HR, skin conductance, and even nonverbal behaviors, such as facial expressions. Given the complexity of human decision-making and the limitations associated with individual sensors, multimodal sensor fusion should increase robustness of the system in detecting relevant states. Here, two or more physiological or behavioral sensors may be combined to gain comprehensive insights into emotion, cognition, and decision-making. Overall, the goal of multimodal sensor fusion is to gather concurrent streams of physiological or behavioral data in real time to create a more comprehensive model of human psychology.

Figure 1 illustrates the motivation to select the subsets of psychological states and physiological and behavioral outputs that can be effectively incorporated into a multimodal sensing and estimation framework. The relationships among these states and outputs are potentially many-to-many, meaning each psychological state is related to possibly many physiological and behavioral features, and any given physiological and behavioral feature is impacted by possibly many psychological states (Cacioppo and Tassinary 1990). Initially, this might seem to be problematic when trying to infer a single psychological quantity from a single modality of output. However, when taking a multimodal approach to inference that captures and accounts for these overlapping relationships, this can actually be favorable and yield a more robust inference of latent psychological states. Here, latent states that correlate to many features are well suited to applying a corroborative inference (Fig. 1, red), while output features that are impacted by many latent states (Fig. 1, blue) are highly informative to that inference and thus particularly valuable to collect. The expected benefit of the multimodal fusion approach is a more robust, convergent mapping from physiological parameters to the psychological state space. This figure also highlights the assumption that there are likely many more physiological and behavioral quantities that can be derived, which inform a relatively compact set of mental states. The relationships between the observable outputs and these latent states are often mediated by complex indirect processes and moderated by interaction effects, and thus are often characterized in the literature by coarse, correlative one-to-one relationships. Therefore, by capturing many-to-many sensor outputs that have such known correlative connections to a common set of states, it should be possible to partially account for this latent complexity and interaction effects, as well as noise, artifacts, and variability.

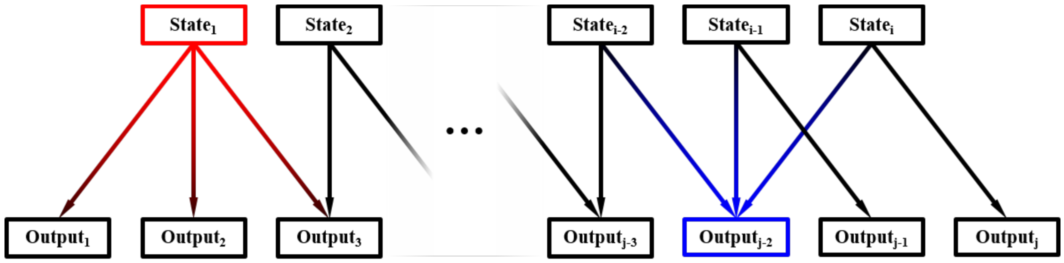

Fig. 1 Diagram illustrating notional connection between latent states and observable outputs

There are substantial bodies of work for specific tasks that target relevant psychological states in relation to one specific modality of physiological measurement, but there has been less emphasis in synthesizing these findings across disciplines. The following subsections review some of the most promising physiological metrics and their known relationships to human states. As such, we propose employing models constrained or informed a priori by these relationships. Here, the latest domain knowledge can describe functional relationships between these states and outputs to maximally inform estimates of latent state when multiple measurement modalities are available.

3.1 Electroencephalography

One of the most widely used metrics in human state estimation research uses EEG techniques and event-related potential analysis to better understand brain dynamics (Prinzel et al. 2003). EEG is a noninvasive technique used to measure the electrical activity of the brain. Although it is subject to many potential sources of noise, filtering methods enable researchers to isolate and process signals from which information relating to task performance and stimulus exposure may be inferred. More importantly, EEG provides information about the intricate brain dynamics that are related to changes in engagement (arousal), motivation, and cognitive workload, with a high time resolution (Rozado and Dunser 2015). Additionally, frequency-based EEG techniques allow researchers to associate changes in specific frequency bands with various cognitive or affective states (Table 1).

Table 1 Common spectral bands associated with EEG measurement and corresponding changes in individual state

EEG band	Typical frequency	Changes in band frequency associated with states and processes
Delta	0.5–3 Hz (amp 20–60 μv)	Fatigue and deep sleep, irregular and slow wave patterns, tumors, and abnormal brain behaviors; present during transition to drowsiness (Lal and Craig 2002)
Theta	4–8 Hz (amp 20–100 μv)	Pleasure/displeasure, drowsiness/fatigue (but not deep sleep); increases have been found in response to demands on working memory (Klimesch 1999); associated with meditative states, and low levels of alertness and decreased information processing (Lal and Craig 2002)
Alpha	8–12 Hz (amp 20–60 μv)	Memory activation and (potential) inhibitory functions (Klimesch 2012); visuospatial attention; appear at eye closure and decrease at eye opening (Tran et al. 2001; Kelly et al. 2006); generally suboptimal performance (*phasic changes*) (Goldman et al. 2002), or selective attention (Foxe and Snyder 2011); responsive to arousal and mental activity (Santamaria and Chiappa 1987; Tran et al. 2001)
Beta (measured as alpha + theta)	13–22/30 Hz (amp 2–20 μv)	Motor control, mental thought and activity, working memory, increased arousal, alertness and engagement, and negative emotion (Güntekin and Başar 2010); sleep spindle frequency ~14 Hz.
Gamma	36–44 Hz (amp 3–5 μv)	Oscillations relate to cognitive processes, consciousness processes (e.g., conscious perception [Melloni et al. 2007; Doesburg et al. 2009]), memory processes (e.g., visual memory [Jerath et al. 2015]) and sudden changes in sensory stimuli. EEG recordings of gamma are highly sensitive to muscle noise (Whitham et al. 2007).

*Some work suggests that gamma can be observed within other frequency bands (~160 Hz) and at a potentially higher resolution in recording technologies besides EEG (e.g., magnetic resonance spectroscopy, see Gaetz et al. [2011]).

For example, within EEG research, the fatigue state has been found to be reliably associated with relative increases in slow wave activity (e.g., delta and theta waves) that are commonly experienced as increases in drowsiness and sleepiness, and are inversely related to cortical arousal (Lal and Craig 2002; Craig and Tran 2012). Cortical arousal has also been linked to the amplitude and frequency of the alpha rhythm (Golan and Neufield 1996), with high-amplitude, low-frequency activity associated with low cortical arousal, and low-amplitude high-frequency activity associated with high cortical arousal (Tran et al. 2001). Additionally, task engagement has also been found to be characterized by increases in beta activity, alpha blocking, and suppression in theta bands (Prinzel et al. 2003). Increases in alpha wave activity have also been found to be associated with decreases in the level of the blood oxygenated signal within the occipital cortex, which in turn produce general performance impairments (Goldman et al. 2002). However, alpha band activity is also associated with selective attention. Attending to a specific region of space, sensory modality, or sensory feature can lead to increases in alpha activity in areas of the brain representing the unattended stimulus properties and a reduction in alpha activity in those representing the attended properties (Foxe and Snyder 2011). Additionally, there appears to be a reliable relationship between changes in alpha and theta bands. More specifically, these frequency bands tend to relate to one another inversely such that increasing task demands (associated with increases in theta) produce decreases in alpha wave activity for example (see Table 2 for full illustration of the alpha–theta relationship). Finally, gamma frequencies have also been observed but it should be noted that they are also difficult to discern due to the fact that the scalp acts as a low-pass filter as well as artifact due to electromyogram (EMG) noise that is generated by facial muscle activity on or near an electrode.

Table 2 Associated changes in alpha and theta power with respect to cognitive performance

	Increases in cognitive performance		**Decreases in cognitive performance**	
	Alpha power	**Theta power**	**Alpha power**	**Theta power**
***Tonic* change**	Increases (+)	Decreases (–)	Decreases (–)	Increases (+)
***Phasic* change (subsecond – seconds)**	Decreases (–)	Increases (+)	Increases (+)	Decreases (–)

Note: Adapted from Klimesch (1999).

If the goal of an adaptive autonomous system is to monitor human states and intervene when suboptimal changes in physiology are detected, it is necessary to understand what effective human performance looks like within a physiological signal. Here, any maladaptive changes within the EEG signal may occur before performance is impacted (Gevins et al. 1990). Within EEG research, qualitatively good performance (e.g., decreases in response time or increases engagement) are characterized by *tonic* increases in alpha and decreases in theta power; the opposite is found if looking at *phasic* event-related changes (see Table 2). Tonic changes refer to sensory inputs that produce event-related potential for the duration of stimulus in a manner, which is gradual during a longer timescale, whereas phasic changes adapt rapidly to a stimulus on a shorter timescale.

Although much research has been dedicated to understanding and documenting the psychological states associated with changes in EEG frequency bands, there are several issues to consider when adopting this methodology (Klimesch 1999). First, EEG is highly susceptible to noise or artifacts that can obscure and make it difficult to interpret the desired brain signals. Moreover, EEG signals can be impacted by several factors including the thickness of the skull and cerebral spinal fluid (CSF) volume, faulty equipment, electrode type and placement, age, baldness, electromagnetic interference (EMI) from nearby electronic equipment (e.g., 60 Hz line noise), muscle activity, and behavior, which produce a) electrical confounds in the signal (e.g., EMG), b) nonlinearities in the measurement system (electrode movement), and c) unwanted neural activity (e.g., something initiated muscle movement/activation [Jackson and Bolger 2014]). Furthermore, processing instructions, focusing on the task, and so on all produce measurable activity in the human brain, but that activity may not be the focus of a specific scientific experiment. Thus, the traditional indicator for quality of EEG data is to have a high signal-to-noise ratio (SNR) (Kappenman and Luck 2010). However, EEG has an inherently poor SNR due to an already small signal that is further attenuated by the issues noted previously. Noise and artifacts are often used interchangeably to describe a feature found in an EEG recording that was not a signal of interest. The term noise, however, is generally used to describe external influences on the signal, whereas the term artifact is more appropriately used to denote activity that is not the signal of interest but a necessary byproduct of the system. For example, eye blinks are a frequent source of artifacts in EEG measurement. Additionally, increases in alpha-band activity might reflect inattentiveness (Goldman et al. 2002), but it could also reflect an increase in selective attention (Foxe and Snyder 2011) necessary for expert performance, making data difficult to interpret. Thus, reported relationships between EEG measures and psychological state must be validated in a relevant task context before they may be considered reliable reporters of a psychological state. These artifacts make it more difficult to assess visually evoked neural activity.

Second, many electrophysiological measures of state might be highly context- and task-dependent. For decades, neural signals have been characterized in impoverished lab environments, where the subjects perform a single, very controlled task to make it possible to elucidate underlying neural dynamics related to the study manipulation. These highly controlled studies are often necessary but make it difficult to translate findings into more realistic scenarios where the stimuli and tasks become much more complex (McDowell et al. 2013). For example, the highly characterized P300 neural waveform has been used to index attention. However, recent studies have found that this waveform is modulated by the state of the person such as whether they are seated or walking (Zink et al. 2016; Bradford et al. 2019; Ladouce et al. 2019). Quality EEG recordings have been made possible by advancements in biopotential measurements systems; however, these EEG recording systems have mostly been optimized for stationary, highly controlled laboratory studies. Recently, EEG methods have been applied outside of laboratory settings in more realistic paradigms where artifacts are increased, further reducing the SNR (McDowell et al. 2013). For example, even allowing face, eye, and head movements can make it difficult to interpret the neural signal. As previously mentioned, sources of artifact include physiological (muscle and eye movements, sweat, etc.), EMI (line noise, other electrical devices, etc.), and movement of the electrode relative to the scalp (may occur due to head movement or whole-body movement of the person [Symeonidou et al. 2018]). Thus, advanced hardware and software solutions to mitigate and correct artifacts during these instances continue to make it possible to interpret neural signals even during whole-body movements, such as slow walking and reaching (Kerick et al. 2009; Bradford et al. 2016). Therefore, the degree to which a paradigm will induce artifacts in the EEG recording should be considered when deciding on including EEG in a sensor suite and choosing the appropriate hardware (Oliveira et al. 2016).

3.2 Electrocardiogram

A second physiological modality used in state estimation assesses changes in autonomic arousal and specifically cardiovascular reactivity. The cardiovascular system is under control of both the sympathetic and parasympathetic branches of the autonomic nervous system (ANS) with changes that have been linked to an individual's level of arousal, anxiety, stress, and fatigue. These changes are primarily captured through an ECG signal, which is measured via electrodes attached to an individual's chest, the surface of the skin, or limbs. Additionally, pulse oximetry, which measures the oxygen level of the blood (i.e., oxygen saturation), can be recorded from the fingertips and may be an important metric to gauge how much oxygen is being carried to the brain and other necessary organs or limbs required for decision-making and effective performance. More specifically, these methods are designed to capture a number of responses including blood pressure, pulse volume, HR, heart period, and heart-rate variability (HRV), to name a few. However, due to the requirements for robust, fieldable state estimation technologies in Army-relevant human–autonomy teaming applications previously mentioned, we focus on the cardiovascular features of HR and HRV, as these appear to best satisfy those requirements. For a complete guide and further reading on all ECG measures mentioned, see the following resources (Task Force on Heart Rate Variability 1996; Berntson et al. 2007).

Before discussing previous research regarding the relationship between cardiovascular reactivity and psychological states, it is imperative to first discuss general physiological processes as well as signal collection methods of the ECG signal. The ECG signal provides information about an individual's cardiac cycle, which can be extracted by examining the typical P, Q, R, S, and T waves of a normal ECG waveform (Fig. 2). The QRS waveform lasts approximately 80–100 ms from the onset of the peak muscle action potential to the ventricular ejection period, which corresponds to the opening of the aortic valve. Further, the QT interval, which can range from 250–500 ms, reflects the time from ventricular excitation to the return resting state. This portion of the ECG signal is dependent upon an individual's HR, where higher HRs correspond to shorter QT intervals (Berntson et al. 2007). Finally, the amplitude of the T wave is sensitive to sympathetic activation; therefore, it has been proposed as a measure of sympathetic control.

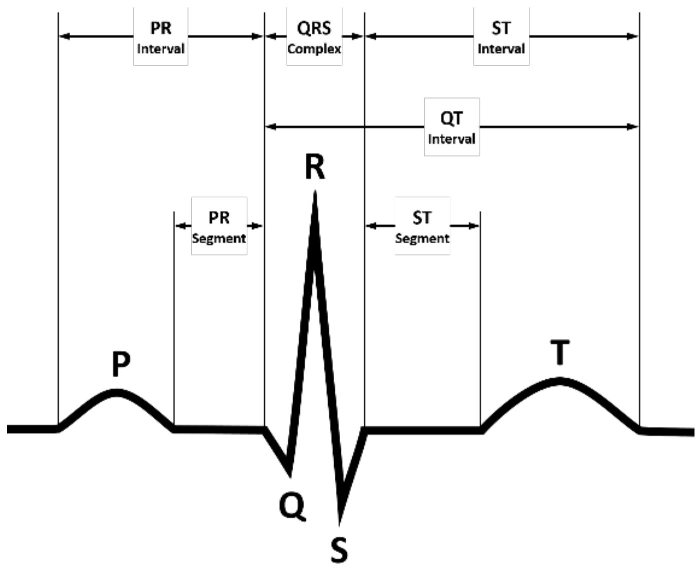

Fig. 2 Diagram of the general structure of the ECG signal, showing the P, Q, R, S, and T components as well as the PR, ST, QT, and QRS intervals (adapted from Berntson et al. [2007])

Although the entirety of the ECG waveform is informative, much of the work in fields, such as affective computing and neuroscience, focus on the "R-spike" within the typical QRS waveform due to the ease of detection. Additionally, the R-spike reflects the important event of electrical depolarization of the ventricles, which allows the aorta to provide blood to vital organs and the brain. The R-spike can be used to compute metrics such as HR (often expressed in terms of beats per minute [BPM]), analyzed via the temporal distance between R-spikes (i.e., the heart period or interbeat interval [IBI]). Note, HR and IBI are reciprocals of one another, meaning IBI can be converted to HR directly (Berntson et al. 2007). Although, these measures are reciprocal, they are not linearly related, as there does appear to be a linear relationship between changes in activity of the parasympathetic and sympathetic branches of the ANS and IBI, but not between the ANS branches and HR (Berntson et al. 2007).

While average HR may convey useful information such as BPM, more specific information about underlying autonomic mechanisms can be attained by looking at how HR varies beat-to-beat, known as HRV. When assessing HRV, two methods are typically employed and relate to *time-* and *frequency*-domain metrics. Time-domain methods reflect the variance among heart periods or IBIs (for further reading, see Task Force on Heart Rate Variability [1996]). One of the easiest-to-measure time-domain metrics is the standard deviation (SD) of the normal-to-normal (NN) beat intervals (SDNN), which is generally derived over a fixed window of time such as 5 min or 24 h. Additionally, other highly inter-correlated, common metrics that are typically derived from HRV assessment include the root mean square of successive differences (RMSSD), the number of interval differences of successive NN intervals greater than 50 ms (NN50), and the proportion derived by dividing NN50 by the total number of NN intervals (pNN50). More specifically, the RMSSD measure is based on the variance of adjacent beat-to-beat IBIs, is independent of baseline heart period, and has subsequently been applied as a measure of high-frequency heart period variability and respiratory sinus arrhythmia (Task Force on Heart Rate Variability 1996; Berntson et al. 2007). When developing a model structure for human state, it is essential to understand typical values associated with such measures. Therefore, Table 3 depicts meta-analytic, average values of commonly used time- and frequency-domain HRV metrics. It can be assumed that values above or below two SDs beyond the average are indicative of a data artifact (Nunan et al. 2010). Additionally, when utilizing ECG metrics, it is important to understand that naturally occurring factors such as respiration, fine motor movements, and even the heart beat itself can produce artifacts in the ECG output; therefore, it may be necessary to combine ECG measures and other physiological modalities to improve artifact attenuation and signal decontamination, discussed in further detail later on.

Table 3 Meta-analytic summary of normal values and range for commonly used measures of HRV

HRV measure	Absolute value				Log-transformed values			
	Mean	SD	Median	Range	Mean	SD	Median	Range
mRR (ms)	926	90	933	785–1,160	n/a	n/a	n/a	n/a
SDNN (ms)	50	16	51	32–93	3.82	0.23	3.71	3.57–4.07
rMSSD (ms)	42	15	42	19–75	3.49	0.26	3.26	3.26–3.41
LF (ms^2)	519	291	458	193–1,009	5.01	1.76	5.02	2.05–7.31
LFnu	52	10	54	30–65	n/a	n/a	n/a	n/a
HF (ms^2)	657	777	385	82–3,630	4.76	1.78	4.96	0.08–6.95
HFnu	40	10	38	16–60	n/a	n/a	n/a	n/a
LF: HF	2.8	2.6	2.1	1.1–11.6	0.69	0.73	0.58	−0.16–1.98

Note: Adapted from Nunan et al. (2010); n/a = not applicable; mRR = mean RR interval; LF = low-frequency spectral power; HF = high-frequency spectral power; LF:HF = ratio of low-frequency power to high-frequency power; nu = normalized unit.

In addition to time-domain measures, ECG research also utilizes frequency-domain (spectral) analyses to assess the degree to which the ECG signal falls within one or more frequency bands (ranges). More specifically, much of the ECG research within this field has used the fast Fourier transform to analyze the HF and LF power spectral density (PSD) to distinguish between different states (Vicente et al. 2016). These methods decompose the overall IBI variance into specific frequency bands, which include high, low, very low, and ultra-low frequencies (see Berntson et al. [1997] and Heart Rate Variability Task Force [1996]). Table 4 depicts these frequency bands along with the typically accepted band range.

Table 4 Average HF and LF spectral band constraint in the ECG

Frequency	Corresponding activity	Average spectral band constraint range
HF	Respiratory sinus arrhythmia/vagal (parasympathetic) control	0.15–0.4 Hz
LF	Sympathetic/parasympathetic* ANS/baroreflex	0.04–0.15 Hz
Very low frequency (VLF)	*	0.003–0.04 Hz
Ultra-low frequency (ULF)	*	0.0–0.003 Hz

Note: * indicates a measure that has yielded mixed results and warrants further investigation.

The HF band corresponds to respiratory sinus arrhythmia (RSA), an individual's breathing cycle, exhibited as an increase in HR during inhalation and a decrease in HR during exhalation. RSA is apparent in both the sympathetic and parasympathetic branches of the ANS and is generally considered to index cardiac vagal control. Consequently, the vagus nerve serves as the main nerve of the parasympathetic nervous system (i.e., vagal nerve activation). Therefore, cardiac vagal control should be viewed as an indicator of how efficiently self-regulatory resources are mobilized and used (Grossman et al. 1991; Berntson et al. 1993, 1994, 1997; Grossman and Kollai 1993; Cacioppo and Berntson 1994). Additionally, the LF band is associated with sympathetic and parasympathetic activity of the ANS and the baroreflex (e.g., the mechanism responsible for controlling acute blood pressure changes via HR, vessel contractility, and peripheral resistance); therefore, anything that enhances sympathetic activation (e.g., exercise, arousal) should also elicit increases in the LF power of the spectral band. However, some claim that the LF also reflects parasympathetic activation as well (Berntson et al. 1997). Note that obtaining accurate measures of the LF bands requires somewhat longer recording times (i.e., a minimum of 4 min or more).

In addition to the actual frequency band, of particular interest within HRV research is the variability of the LF/HF ratio. The ratio is considered to be indicative of sympathetic and parasympathetic autonomic balance. However, this finding is somewhat controversial and warrants further research (Eckberg 1997, 1998; Heathers 2014; Quintana and Heathers 2014). Here, it is assumed that the variability in the LF band is driven by both branches of the ANS. Therefore, an increase in sympathetic control would increase LF but not HF variability, thus reducing the index value (Berntson et al. 2007). Typically, PSD techniques support the use of the LF/HF ratio as a measure of drowsiness, where drowsiness without stress is typically reflected as high HF and low LF (i.e., a low LF/HF ratio) (Patel et al. 2011). Additionally, other studies assessing sleep-deprived subjects showed that as subjects became sleepy their LF component decreased, while the HF component increased (e.g., the LF/HF ratio decreased when subjects became sleepy [Vanlalchaka and Zonunmawii 2018]). However, there is some debate around utilizing this measure, as LF may uniquely reflect sympathetic regulation. The PSD band has also been further subdivided to include VLF and ULF. However, VLF and ULF are of debated origin and have not received much attention within psychophysiological research, as both stem from autonomic branches and are potentially influenced by slower endocrine influences related to temperature regulation and other basic homeostatic regulatory functions. There has been some work investigating the relationship between the "mid-frequency" band as a way to quantify mental workload and the baroreceptor function (Boucsein and Backs 2000; Van Roon et al. 2004).

Within HRV research, much work has focused on changes in HRV as a reliable predictor of emotion regulation and suppression. Typical findings tend to show that increases in HRV have been associated with increases in self-regulatory effort and emotion regulation, increased attentional control, and behavioral flexibility (Segerstrom and Nes 2007). Typically, an inverse relationship between HR and HRV is observed, as they relate to changes in human state (Table 5). For example, an increase in HR indicates a generally arousing state such as a stress or anxiety response, whereas an increase in HRV indicates a high fatigue state or increased emotion regulation (Wohlebler et al. 2018). In fact, O'Hanlon (1972) found that HRV increased in response to time on task, an indicator of a fatigue state, and consequently resulted in decreased performance and an apparent vigilance decrement, but then decreased when sudden events alerted drivers and gained their attention, also resulting in a slight increase in their HR.

Table 5 Directional relationship between cardiovascular measures and emotional state variables

	Individual state or process				
Measure	**Arousal**	**Emotion regulation**	**Fatigue**	**Stress**	**Trust**
HR	Increase (+)	*	Decrease (-)	Increase (+)	Decrease (-)
HRV	Decrease (-)	Increase (+)	Increase (+)	Decrease (-)	Increase (+)

Note: * indicates that a relationship is largely unknown or warrants further investigation.

There are some key caveats associated with this particular modality. Mainly, the ECG signal does appear to reflect changes in RSA, which is also influenced by factors such as posture, movement, age, respiratory depth, and general fitness level of the individual. Therefore, it has been recommended that, at a minimum, respiratory measures be recorded and entered as a potential confound or covariate during analysis (Berntson et al. 2007). Additionally, the amount of cardiac change seen during an experimental manipulation can differ depending on the chosen metric, experimental manipulation, and baseline differences among individuals. Therefore, using the IBI measure, rather than HR, is recommended to more accurately interpret the data because it is not as susceptible to these influences (Berntson et al. 1995).

3.3 Electrodermal Activity

Psychophysiological activation can also be measured specifically through changes in electrodermal activity (EDA). According to Woodworth and Schlosberg (1954), EDA is "perhaps the most widely used index of activation" (p.137) and is essentially a sensitive, peripheral index of sympathetic nervous system activation. Measures of EDA monitor the skin's electrical activity by measuring the amount of sweat that is secreted from the sweat glands that is generated by physiological or emotional arousal when individuals are exposed to emotionally evocative stimuli. Here, an increase in sweat results in higher skin conductivity, which is typically expressed in units of microSiemens (μS). However, before continuing the discussion of EDA measures and their relation to psychophysiology, it is useful to outline the components of the EDA signal and appropriate metrics for assessment.

The EDA signal includes 1) a general *tonic level*, which relates to the slower-acting background characteristics of the signal (referred to as skin conductance level [SCL]) and 2) the more specific and rapid *phasic* component (also referred to as a skin conductance response [SCR]), which results from sympathetic activity. Generally, the tonic level reflects changes in autonomic activation and general states of arousal or alertness, which are illustrated via the overall level of slow increases and decreases in skin conductance signal over time. The faster phasic component reflects the magnitude or degree of activation to a specific stimulus such as attentional processes and individual differences. However, this portion of the signal only makes up a small portion of the EDA complex. It appears that both signals are important and may rely on different neural mechanisms for activation (Nagai et al. 2004; Dawson et al. 2017).

Typically, EDA methodology involves continuous measurement of the SCL, which tends to gradually decrease while individuals are at rest, rapidly increases when a novel stimulus is introduced, and then gradually decreases again when subjects are at rest or experience habituation (Dawson et al. 2017). Additionally, the presentation of a novel, personally significant, unexpected, or aversive stimulus will likely result in a SCR, referred to as a "specific" SCR. Conversely, if a SCR occurs in the absence of a stimulus, it is then referred to as a "spontaneous" or "nonspecific response" (NS-SCR). SCRs occur after an event has happened; therefore, the sampling window following the presentation of a stimulus is generally between 1–4 s to allow for the accurate capture of the response. Therefore, an SCR that begins between 1 and 4 s following stimulus onset should be considered a response to that stimulus; however, it is methodologically important to be sure to select these fairly short latency windows to reduce the likelihood that a NS-SCR be taken as a SCR (Dawson et al. 2017). Figure 3 illustrates the principal EDA components.

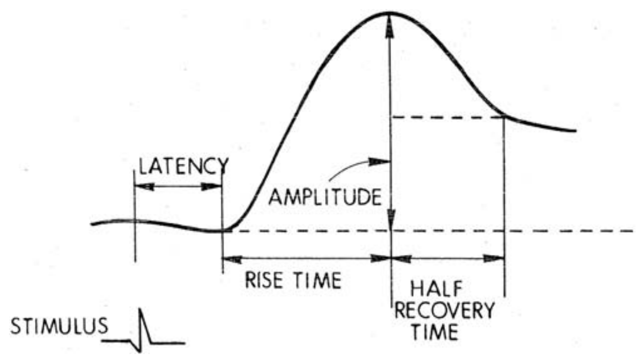

Fig. 3 Graphical representation of principal EDA components (Dawson et al. 2017)

When assessing EDA, a number of metrics exist. The most commonly reported measure is the size of the SCR, which reflects the increase in conductance from the onset of a SCR to its peak and illustrates the degree of sympathetic activation. Additionally, when measuring the various components of the EDA complex, it may be useful to have a reference for specific EDA measures and a typical range of responses, which are outlined in Table 6.

Table 6 Electrodermal measure, definition and typical range value

Measure	Definition	Typical values
SCL	Tonic level of electrical conductivity of skin	2–20 µS; 4–5 µS at rest
Change in SCL	Gradual changes in SCL measured at two or more points in time	1–3 µS
Frequency of NS-SCR	Number of SCRs in absence of identifiable eliciting stimulus	1–3 per minute
SCR amplitude	Phasic increase in conductance following a stimulus onset	0.1–1.0 µS
SCR latency	Temporal interval between stimulus onset and SCR initiation	1–3 s
SCR rise time	Temporal interval between SCR initiation and SCR peak	1–3 s
SCR half recovery time	Temporal interval between SCR peak and point of 50% recovery of SCR amplitude	2–10 s
SCR habituation (trials to habituation)	Number of stimulus presentations before two or three trials with no response	2–8 stimulus presentations
SCR habituation (slope)	Rate of change of ER-SCR amplitude	0.01–0.5 µS per trial

Note: Adapted from Dawson et al. (2017).

While EDA is sensitive to a wide variety of stimuli, for our work, it cannot be directly linked to one particular psychological process (Landis 1930). For example, it is nearly impossible to discern a SCR as an “attentional”, “anxiety”, or “anger” response. However, the value of SCRs become psychologically meaningful or interpretable when one takes into account the condition or experimental paradigm in which the SCR occurred. Here, conclusive interpretations must be made within tightly controlled experimental settings, or with sufficiently complete capture of environmental stimuli that might produce these responses.

Within a generalized, experimental setting, tonic EDA has been linked to energy regulation or mobilization processes. These can be interpreted via information processing theories, which propose that tasks requiring an effortful allocation of attentional resources are associated with heightened autonomic activation and in turn larger tonic EDA fluctuations (Jennings 1986). A further view proposes that the stress associated with performing laboratory tasks leads to increased sympathetic activation and EDA arousal, not necessarily activation associated with attentional resource allocation (Dawson et al. 2017). Additionally, SCRs have been reliably found to relate to various arousal dimensions. For example, SCR magnitude has been found to increase in conjunction with subjective arousal ratings of positively valenced pictures (greater in response to erotic pictures compared to pictures of beautiful flowers) and negatively valenced pictures (greater for snakes than for tombs in a cemetery). However, such SCRs tend to be most sensitive to stimulus novelty, intensity, and personal significance (Lang et al. 1993; Cuthbert et al. 1996). Consequently, EDA measures have also been used within medical settings to identify precursors to impending psychotic relapses (Hazlett et al. 1997). Here, it has been hypothesized that sustained and heightened sympathetic activation can interfere with efficient cognitive processing capacity (Nuechterlein and Dawson 1984).

In addition to the previously mentioned findings associated with EDA research, it is also vital to briefly mention one individual difference factor associated with changes in tonic and phasic EDA. Most researchers have concluded that EDA can be viewed as a relatively stable trait of the individual, though one that is subject to individual differences, which may further impact specific responses to environmental stimuli. Generally, individuals can be categorized as electrodermal “labiles”, who show high rates of NS-SCRs and slow SCR habituation, while electrodermal “stabiles” are those who show relatively few NS-SCRs and fast SCR habituation. More specifically, electrodermal lability has been found to be reliable over time, with labiles outperforming stabiles on sustained attention tasks, typically marked by a deterioration of correct target detections over time (Davies and Parasuraman 1982), which may reflect their increased ability to maintain attentional focus (Crider and Augenbraun 1975; Hastrup 1979; Vossel and Rossman 1984; Munro et al. 1987). Conversely, several studies have confirmed that a vigilance decrement tends to be more pronounced among the EDA (specifically SCR and habituation rate) of stabiles (Munro et al. 1987; Koelega et al. 1990). Therefore, in the words of Katkin (1975, p. 172) "electrodermal activity is a personality variable that reflects individual differences in higher central processes involved in attending to and processing information".

Finally, when discussing each modality, it is crucial to also present the caveats or disadvantages of each. First, EDA is a relatively slow-moving response system, with SCR response latencies between 1–4 s. Therefore, rapidly occurring processes may not be measurable via this modality. Second, it is vital to allow enough time to pass between stimulus presentations as a SCR needs time to "recover" following activation. If an appropriate amount of time is not allowed between stimulus presentations, subsequent SCRs may simply be superimposed on the recovery phase of the first response. For example, the amplitude or size of a subsequent SCR may be smaller given its occurrence immediately following a previous SCR, although the degree of distortion of the subsequent SCR is a function of the size and time of the first response (Grings and Schell 1969). Third, as previously mentioned, there may be large-scale variability due to individual difference factors such as electrodermal lability. Therefore, it is recommended that researchers use within-subject standardized scores (Ben-Shakhar 1985). Finally, as with most physiological modalities, response artifacts may be elicited by movement or deep breaths; therefore, it is vital to record these as well to document which responses are true SCRs and which are NS-SCRs.

Despite these limitations, this particular physiological modality provides valuable insight into the lower-level, automatic and subconscious arousal that a person experiences when exposed to emotionally loaded or novel stimuli. As such, EDA is unique in that it reflects relatively direct and undiluted representations of sympathetic activity. Therefore, increases in SCL or SCRs reflect increased tonic and phasic sympathetic activation. This is in direct contrast to other physiological modalities that measure other ANS functions (e.g., pupil diameter, HR, and so on) as changes in response to various stimuli cannot be directly linked to sympathetic or parasympathetic activity. In other words, responses may be due to one or both systems working in conjunction with each other and cannot be isolated. Therefore, if one wishes to measure direct sympathetic activation, EDA measurement is best, though a broader representation of both sympathetic and parasympathetic activity requires other measurement modalities such as HR.

3.4 Eye Tracking

Another modality for gauging cognition, arousal, and fatigue is eye tracking. Eye tracking offers insight into overt visual attention by monitoring where individuals direct their eye movements at certain points in time, as well as dilation of the pupil, which is tightly linked to activity in the ANS (Joshi et al. 2016). Pupil size is constantly in flux, reacting to environmental changes in luminance via constriction (in response to brightness) and dilation (in response to darkness). These unconscious responses help to optimize visual acuity and sensitivity for rapidly changing visual scenes. However, even in cases where luminance is held constant (e.g., in controlled laboratory settings), there is a rich literature documenting psychosensory influences on pupil size (Beatty 1982) including significant effects related to arousal (Loewenfeld and Loewenstein 1993), sleepiness (Wilhelm et al. 1998), attention (Gabay et al. 2011), emotion (Bradley et al. 2008), control state (Jepma and Nieuwenhuis 2011), and mental processing load (Kahneman and Beatty 1966; Granholm et al. 1996). Table 7 provides a selection of pupil features and their relationship to psychological states.

Since pupil size is strongly linked to the ANS (via the locus coeruleus-norepinephrine system), eye-tracking metrics can also be useful for validating and complementing other autonomic-linked signals derived from EDA and ECG measurements. Of note, pupil dilation occurs on a much slower timescale than does EEG (on the order of seconds versus milliseconds, respectively). Therefore, combining EEG with pupillometry measurements could improve continuous monitoring of fatigue, arousal, and mental engagement when individuals are exposed to emotional or cognitively demanding tasks.

Table 7 Relationship between specific pupil features/metrics and psychological states

Pupil feature	Definition/ calculation	Range	Psychological state	Relationship	References
Baseline pupil size	Mean pupil size to calculate the size of the pupil over a period of time (usually on the order of seconds) prior to or during a task.	2.0–8.0 mm	General arousal; control state (exploration vs exploitation)	Linear (larger pupil size = higher arousal); Curvilinear (inverted "U" shape)	(Loewenfeld and Loewenstein 1993; Jepma and Nieuwenhuis 2011)
PUI	The pupil's tendency to instability; defined by the sum of absolute changes in pupil diameter (in mm) based on a sample frequency of 1.5625 Hz (see ref for formula).	2–12 mm	Sleepiness/ fatigue	Linear (larger PUI = more sleepiness)	(Leudtke et al. 1998)
Index of cognitive activity (ICA)/index of pupillary activity (IPA)	Wavelet analysis decomposes raw pupil signals to distinguish the light reflex from cognitive responses. Frequency of rapid super-threshold pupil dilations indicate cognitive activity (see refs for formula)	0–1 Hz	Index of cognitive activity and workload	Linear (larger ICA/IPA = higher cognitive load)	(Granholm et al. 1996; Marshall 2000; Duchowski et al. 2018; Vogels et al. 2018)
Peak pupil dilation	Maximum amount of pupil dilation in a short time window following sensory stimulation	0.1–1.0 mm	Task-Induced Mental Load	Linear (larger amplitude = higher mental load)	(Kahneman and Beatty 1966)
Peak pupil latency	The amount of time required to reach peak pupil dilation	500–3,000 ms	Decision Making/Execution	Linear (longer latency = longer decision process)	(Cohen Hoffing et al. 2020)

Note: Specific formulas for each pupil feature can be found in the associated reference(s).

To the extent that pupil dilation provides a proportional index of psychological states (e.g., arousal, engagement, cognitive load, and so on) during task execution, pupillometry could provide an effective, noninvasive tool for measuring changes in human state, particularly if confounds due to changes in luminance can be accounted for or controlled. There have been various attempts to model the influence of luminance on pupil size, with a majority of work focused on predicting stable-state changes in baseline pupil size as a function of luminance (for review, see Watson and Yellott [2012]). While these models are very good at capturing expected changes in mean pupil size following relatively long periods of adaptation (greater than 10 s) to well-controlled luminance levels in the laboratory, they are not designed to predict the rapid patterns of pupillary change when a person is confronted with complex and dynamic naturalistic scenes. More recent work by Korn and Bach (2016) proposed a method to model pupillary dynamics in response to slightly more rapid luminance changes (5-s stimuli), expressed as the sum of two linear time-invariant systems that capture the differential time course of dilations and constrictions using gamma response functions. Despite the success of this and related approaches (Denison et al. 2020), there is still no widely accepted model to date that can accurately predict luminance-based changes in pupil size when viewing complex visual scenes (e.g., movies) or when a person is situated in a real-world environment. This challenge will likely need to be solved before the full potential of pupillometry is realized for psychological state monitoring in real-world applications (Cohen Hoffing et al. 2020).

In addition to pupil-based measures of human state, eye-movement metrics provide a type of trace of where an individual directs their eye gaze. This is important because *what* a person is looking at is thought to reflect the cognitive symbol currently being processed (Just and Carpenter 1976). Therefore, it is possible to identify stimuli that are important or emotionally salient to the subject. Eye-tracking data, then, are useful to trust research, for example, because eye-movement metrics provide indications of what information the subject might find important for trust calibration. For example, repeated gaze fixation on an area of interest, such as one display compared to another, would allow an inference that the subject finds the display useful or critical to the current task (Just and Carpenter 1976; Poole and Ball 2006), whereas increased fixation duration might indicate that the subject finds the information difficult to understand (Poole and Ball 2006). Further, the pattern and targets of gaze fixation may indicate the direction of an impending decision (Pool and Ball 2006; Glaholt and Reingold 2011; Gildöf et al. 2013) and an estimate of the level of cognitive effort being expended while processing information (Marshall 2007). Typical eye movements consist of quick and frequent movements called saccades, which are interspersed with periods of steady gaze fixations (Poole and Ball 2006). Fixations are valuable metrics as their timing can indicate both perception and cognitive activity. For example, fixations that last for longer periods of time can indicate that a person is having difficulty extracting or comprehending their visual information. More specifically, fixations associated with cognitive processing (i.e., fixations lasting between 150–900 ms) tend to decrease as a person struggles to maintain focus and attention, a state associated with the onset of fatigue. However, some have argued that mean fixation duration may relate more strongly to workload than fatigue (Poole and Ball 2006). Table 8 provides typical values associated with fixation and fatigue. As with SCR features of EDA, the use of eye-gaze measurements to infer psychological states and processes requires sufficient capture of environmental stimuli that those measures correspond to.

Table 8 Eye-tracking fixation periods (in ms) that are associated with human state

Fixation (in ms)	Associated human state
≤150	Low level unconscious control but not deep processing
150–900	Cognitive processing
150–900 or longer	Difficulty maintaining focus and attention, onset of fatigue
≥900	Indicative of staring and minimal visual sensory processing

Note: Adapted from Schleicher et al. (2008).

Additionally, the PERCLOS measure has been widely used in human state-estimation research (Wierwille et al. 1994). PERCLOS is defined as the proportion of time that a person's eyes are more than 80% closed, though other research has used a 70% cutoff value (Dinges et al. 1998). Increased eye closures are typically associated with decreased arousal or increased fatigue states; therefore, the PERCLOS measure is a noninvasive way to allow researchers to gauge such changes.

3.5 Facial Expressions

One of the strongest visually observable indicators for emotion is the human face. We can read emotions in others based on even minute changes in the posture of prominent or key facial features such as the eyes, brows, lids, nostrils, and lips. The human face includes over 40 structurally and functionally autonomous muscles, each of which can be triggered independently of each other, though they are innervated by a single nerve, referred to as the facial nerve, which emerges from within the brainstem and branches off to all muscles. Facial muscle activity is highly specialized for expression and allows us to share social information with others to communicate both verbally and nonverbally. Humans can produce thousands of expression variations; however, there is only a small set of distinct facial configurations that we associate with certain emotions, irrespective of gender, age, cultural background, and socialization history (to an extent), which include the emotions happiness, anger, surprise, fear, contempt, sadness, and disgust.

Facial expressivity has been shown to be related to appraisal and coping mechanisms, as well as stress, fatigue, and trust. As such, computer-based facial expression analysis attempts to mimic human coding of emotion as it captures raw, possibly unfiltered emotional responses to engaging stimuli. Additionally, past studies have found that automatic computations of facial expressivity are comparable to manual annotations of emotional expressions (Neubauer et al. 2017b) and have been utilized in a number of both clinical and experimental studies (Batrinca et al. 2013; DeVault et al. 2014; Chollet et al. 2015; Venek et al. 2016; Neubauer et al. 2017a; Parra et al. 2017). Therefore, this modality provides evidence that automatic behavior trackers have the ability to support clinical assessments and provide researchers with much needed objective assessments of behavioral indicators of stress, trust, or even team cohesion (Neubauer et al. 2020).

Facial expressions can typically be assessed with two different methods. First, tracking of facial electromyographic activity (fEMG) records the activity of facial muscles with electrodes attached to the skin surface. fEMG detects and amplifies the electrical impulse generated by the respective muscle fibers during contraction. For example, the right/left corrugator supercilii (e.g., eyebrow wrinkler) is a small, narrow, pyramidal muscle near the eyebrow, generally associated with frowning. The corrugator draws the eyebrow downward and toward the face center, producing a vertical wrinkling of the forehead. This muscle group is active to prevent high sun glare or when expressing negative emotions such as suffering. In addition, the right/left zygomaticus is a muscle that extends from each cheekbone to the corners of the mouth and draws the angle of the mouth up and out, typically associated with smiling. Therefore, when facial expressions are apparent so too are the associated electrical impulses that result from movement.

The second method involves visual observation of facial landmarks. This can be assessed by either manual coding, using the Facial Action Coding System (FACS; Ekman and Friesen 1978), or automated, computer vision-based detection of facial activity. The FACS represents a standardized classification system of facial expressions for expert human coders based on anatomic features. Coders examine videos of an individual's face and describe any occurrence of facial expressions as combinations of elementary components called Action Units (AUs). Each AU corresponds to an individual face muscle or muscle group and is identified by a number (AU1, AU2, etc.). Figure 4 shows the location on the face of various AUs. All facial expressions can be broken down into their constituent AUs. In other words, facial expressions can be likened to "words", while AUs are the "letters" that make up those words. Table 9 illustrates which AUs can be calculated to reveal changes in the universal emotions described previously, and Fig. 4 depicts a diagram of the facial regions affected by those relevant AUs. The alternative method for visually coding the activity of the face utilizes computer-vision algorithms to automatically detect a human face and employ feature detection to identify facial landmarks such as eyes and eye corners, brows, mouth corners, and nose tip, and so on. With the feature detection, an internal face model is adjusted in position, size, and scale to match the individual's actual face. Whenever the respondent's face moves or changes expressions, the face model adapts and tracks. Feature classification then translates the position of landmark facial features into AU codes, and thus emotional states and other affective metrics (e.g., OpenFace [Baltrušaitis et al. 2016]).

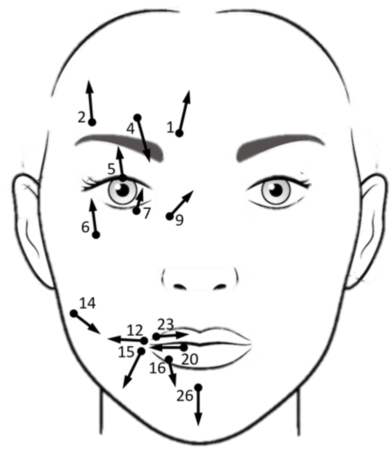

Fig. 4 Diagram of facial expression AU locations and direction of change (right half of face only)

Table 9 Facial expression emotion calculation from single AUs (Ekman and Friesen 1978)

Emotion classification	AUs
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

Most of the published research on computer vision application to detect human state have focused on fatigue assessment and typically relied on analyses focused on eye tracking and head movements (Gu and Ji 2004; Zhang and Zhang 2006; Dong et al. 2011). Here, we argue that facial expressivity metrics can also be used to measure changes in affect, and specifically *affect-based trust*, an emergent attitudinal state in which the individual makes attributions about the motives of the automation (McAllister 1995; Burke et al. 2007). Analyses using these features may be important to human–autonomy teams because these data (e.g., emotions, body postures, and facial expressions) can provide insights into behavioral patterns that have been linked to affiliation, empathy, and assessments of team member trustworthiness. This line of research is critical because it will be necessary to develop autonomous systems that can robustly perceive and respond to affective changes of teammates if human–agent teams are to be successful (Bartlett et al. 2004).

This presents an interesting case because human–human teams may communicate nonverbally through changes in emotional expression (i.e., we regularly gather information from our partners' faces). For example, if an individual is worried about a particular decision they made or need to make, they may seek confirmation or alternate solutions from nonverbal features of their partner. Alternatively, if something negative impacted the state of the team and one team member responds appropriately (e.g., some sort of negative affective response) whereas the other does not (e.g., they smile in response to a team failure), then trust and eventual cohesion may suffer. Within the human–autonomy team domain, it is important to acknowledge that humans may not get the same kind of nonverbal feedback that they normally would from a human equivalent. Given scenarios such as these, additional considerations should be made with regard to the communicative design of human–autonomy teams. To date, collection of this data modality has been exploratory within our lab, though we anticipate that facial expression measurements will provide corroborative support for other relevant behavioral and physiological measures that indicate changes in emotional state.

4. Important Psychological States

Thus far, we have assembled information on several methods to extract data from humans via noninvasive or wearable physiological and behavioral sensors as they inform typical baseline levels and relevant changes in various latent states mentioned previously. Although, there are many states of interest within human–autonomy teaming, for this report, we focus on the states associated with fatigue, stress, trust, workload, and vigilance, as these are some of the most extensively researched topic areas and are believed to provide the most valuable input for the types of modeling previously discussed, given the level of overlap onto the set of sensing modalities described previously. The following sections outline these states individually and include information from the literature regarding the typically known relationships that exist between latent states and observable physiological and behavioral outputs. The purpose of this section of the report is to provide a concise summary of these states, such that researchers aiming to perform real-time inference of a particular state, or set of states, can more easily select sensing modalities and apply the relevant domain to that inference.

4.1 Stress

Stress has been defined as “the force that degrades (task) performance capability” (Hancock and Warm 1989). Task performance is frequently stressful, as evidenced by laboratory and field research, such as vehicle driving, industrial work, and military operations (Matthews et al. 2000a). Tasks may be intrinsically demanding, because they impose high workload or time pressure, or have a high likelihood of failure. The environmental context in which the task is performed may also be a source of stress. Operational settings may be noisy, hot, or dangerous; require prolonged, fatiguing work shifts; or some combination of these. Social factors such as interactions between team members may also elevate task demands and, in turn, stress. Task-related stress may have a variety of consequences including acute emotional responses, performance impairments, and long-term impacts on health and well-being.

Stress states are sensitive to both external influences of the environment (e.g., noise), task demands, and individual competency in managing those demands. States, in turn, influence information-processing characteristics, which are expressed in observed performance. For example, stress can interfere with an individual’s ability to maintain focus and apply effortful regulation during a task, which can result in a degradation of individual and team performance. Thus, stress states have both physiological and psychological aspects (Fairclough and Venables 2006). From a psychological standpoint, the leading theory centers on the transactional theory of stress and emotion (Lazarus 1999), which views the stress response as a twofold transaction comprising 1) individual *appraisal* of the task and 2) individual *coping* mechanism. This “transaction” is a subjective experience created by the individual to apply meaning to a current encounter. Therefore, a key insight concerning stress and performance is that individuals actively regulate their handling of task demands in stressful environments, for example, by varying effort or strategy. A further insight is that the task itself is often a source of stress, especially when it is appraised as taxing or exceeding the individual’s competence. Hence, both performance and well-being may depend on the coping strategies adopted by the individual in response to task demands. Stress theory also emphasizes the importance of individual differences in coping, where the choice of coping strategy depends on the individual’s appraisals of environmental demands and of their own personal abilities to execute effective coping mechanisms.

4.1.1 Stress Measurement

One way to investigate task stress is to focus on changes in the individual's mental state. For example, stress may be accompanied by negative emotions such as anxiety, anger, and unhappiness. Emotions, in this context, are a structured set of multiple psychological processes, including somatic responses, subjective feelings, processing biases, and action tendencies that serve a functional purpose. For example, the various components of fear promote awareness of danger and readiness for escape. An emotional *state* is thus a temporary configuration of multiple processes that may produce a variety of behavioral changes. Operationally, stress states may be assessed through a number of self-report measures such as the Dundee Stress State Questionnaire (see Matthews et al. [2012b] for further reading), or through psychophysiological response.

4.1.2 Observable Correlates of Stress

Generally speaking, stress can be viewed as an arousing state that has a negative-valence component. It is a state that is also associated with changes in anxiety. Therefore, much research on the psychophysiological correlates of stress have also shown stress to coincide with changes in arousal or anxiety. In fact, early studies of mood showed that both energetic arousal (similar to engagement) and tense arousal (similar to distress) correlated with various measures of autonomic arousal, including HR and EDA (Thayer 1978). In other words, activation of the ANS, shown through an increase in HR, should indicate a generally arousing scenario such as a stress or anxiety response. However, as noted earlier, some physiological modalities (e.g., EDA) cannot explicitly delineate changes in stress, as they can only measure changes in arousal, but not necessarily the valence associated with that arousal.

When faced with task demands, an individual's level of arousal should increase. However, the valence associated with that arousal falls within one of two categories: a threatening "stress state" or an engaging "challenge state" (Tomaka et al. 1997). A psychological "challenge" occurs when individuals believe they possess the cognitive ability to meet required task demands. This manifests as psychological "threat", and in turn a distressing stress state, when they believe that task demands outweigh their cognitive or physical resources. States of psychological "threat" are indexed via an increase in vasoconstriction (e.g., total peripheral resistance [TPR]) and a decrease in cardiac output (CO). TPR reflects vasodilation (increased blood flow) and vasoconstriction (decreased blow flow), which are related to parasympathetic and sympathetic activity. CO is the amount of blood pumped in terms of volume per unit time. It has been shown that TPR unambiguously increases in a threat state and decreases in a challenge state, whereas CO either remains unchanged or decreases in a threat state and increases in a challenge state (Tomaka et al. 1997; Neubauer et al. 2017a).

Additionally, stress states can also vary temporally. This temporal variation is also associated with changes in the heart and various blood vessels. For example, acute stress, or momentary/short-term stress, is associated with an increase in HR and stronger contractions of the heart muscle. This aspect of stress is also associated with a flood of stress-related hormones, which are designed to prime the body for action. Furthermore, blood vessels, which are responsible for transporting blood to large muscles, and the heart dilate, thereby allowing more blood and thus more oxygen to these areas. This state is also associated with an increase in blood pressure; however, it is a temporary state and once the episode is over the body returns to equilibrium.

4.2 Workload

Mental workload has been defined as the ratio of task demand to allocated resources (Wickens 2002). More specifically, workload has been defined as "the relation between the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator" (Parasuraman et al. 2008, pp. 145–146) and has been shown to be a critical factor that influences team effectiveness in human–autonomy teams. The foundation of this research stems from the automated or driverless vehicle literature, which suggested that the inclusion of automation aides help reduce mental workload and positively benefit performance (Scribner and Dahn 2008; Yang et al. 2009; Scribner et al. 2013). Within this area of research, one of the primary purposes of automation is to reduce and regulate workload, so that individual levels of workload fall within a "moderate" level and hence avoid task overload and underload. Thus, automation implementation has been successful for both mental workload (Wiegmann et al. 2001) and in the implementation of action (Yang et al. 2009).

However, when it comes to measuring workload, especially in human–autonomy teams, it is often difficult to separate workload from other constructs such as stress, trust, and situation awareness. Specifically, research has suggested that stress and workload are variable and unpredictable. More specifically, unregulated increases in stress and workload can lead to degradations in trust (Biros et al. 2004; Cosenzo et al. 2006; Wang et al. 2009). Further, increases in mental workload associated with autonomous systems can lead to degradations of trust (e.g., in combat identification tasks [Wang et al. 2011]). Similarly, trust in vehicle control systems is higher when workload is low (Spain and Bliss 2008), and the system is perceived to be more useful (Donmez et al. 2006). This relationship is mediated by situation awareness. While the relationship among workload, situation awareness, and autonomy is complex, overall, increased automation leads to reduced workload. However, when there is an imbalance between situation awareness and workload, there is often an increase in performance-related errors (Beer et al. 2014). In other words, if increased automation reduces workload and in turn situation awareness, then an "out of the loop" performance problem may occur (e.g., individuals become passive rather than active participants in a task) and mistakes, errors, and degradations of trust can be common.

4.2.1 Workload Measurement

Workload has a longstanding measurement history. The most prominent means of measurement is through the NASA Task Load Index (NASA-TLX; Hart and Staveland, 1988). This self-report questionnaire measure provides workload assessment specific to mental demand, physical demand, temporal demand, performance, effort, and frustration. According to Hart and Staveland (1988), workload can further be defined as "the cost incurred by a human operator to achieve a particular level of performance" (p. 2). While it has been used in a variety of settings, its applicability for adequate assessment of workload when intelligent agents are part of the team is unknown. While general findings suggest that workload decreases as autonomy increases (e.g., less use of teleoperation), self-report only provides a portion of the outcomes of workload. Thus, Steinfeld and colleagues (2006) identified a critical need to identify nonintrusive measures of workload to characterize the human team member in real time. They went on to suggest a set of behavioral indicators of workload related to the rate of interventions by the human, ratio of time each agent (human or artificial) spends performing the task, and the number of artificial agents each human can control (known as "fan out").

According to Goodrich and Oleson (2003), fan out is a measure of how many similar robots can be controlled by a single person. It can be used as an indicator for robot hand-offs and provides an upper limit of workload. Building on the research by Scholtz et al. (2003), the number of interventions per unit time, along with average time needed for the intervention, and the effectiveness of the intervention may be an indicator a workload (however, this may also be indicative of trust). Additionally, the ratio of operator time to robot time on task builds on the research from Yanco and colleagues (2004), which can be used to identify the balance of workload among team members, human or autonomous. More recent research has shown that there is a divergence of self-report measures with physiological and performance-based measures. Here, Matthews et al. (2020b) suggest three possible solutions to this divergence by creating reliable and valid measurement, selecting appropriate measures based on the predictability of a specific outcome, and selecting representational workload measurements that correspond to real empirical phenomena, though current measurement techniques seem to measure different constructs.

4.2.2 Observable Correlates of Workload

Research has pointed to a number of observable correlates of workload, including HR, HRV, EDA, eye metrics, and EEG. Typically, within workload research, patterns show that HR increases and HRV decreases as workload increases (Delliaux et al. 2019). More specifically, high mental workload tends to raise HR and blood pressure (Schnall et al. 1990; 1994; Wilson 1992, Veltman and Gaillard 1996; Hjortskov et al. 2004), while lowering HRV (Mulder and Mulder 1981; Hjortskov et al. 2004), which is likely due to increased sympathetic activation or decreased parasympathetic activation. Moreover, few studies have looked at the link between mental workload and the arterial baroreflex, which is a critical component of the ANS sympatho-vagal balance. This reflex seems to be impaired by increased mental workload and is thus identified as a significant cardiovascular risk factor (Mulder and Mulder 1981). Other research has also identified significant relationships between increased cardiac activity and changes in specific EEG band frequencies (e.g., alpha and delta bands), which were seen in response to various task-related demands that varied in their level of workload (Wilson et al. 2002).

Additionally, Wilson et al. (2002) argued that HRV is a less sensitive measure of workload than HR. They also found that blink rates decreased during more visually demanding (i.e., higher workload) segments of a task. Workload has also been associated with changes in EDA, where tonic EDA increases in response to increased workload demands, while phasic EDA responses tend to be more frequent. Finally, blood pressure and pupil diameter have also been found to increase in conjunction with workload (Hughes et al. 2019). For example, a classic study by Hess and Polt (1964) found that the pupil response was a good indicator for mental activity by finding that the size of the pupil increases in proportion to task difficulty and workload (see also Beatty [1982]). Additionally, they argued that changes in the pupil, in response to changes in mental workload, occur at short latencies at the onset of processing and subside quickly after processing is complete (i.e., latency onset between 100 and 200 ms) (Beatty 1982). This work led Kahneman (1973) to rely on what is known as the "task-evoked pupillary response" as a primary determinant of processing, and hence workload, for his classic theory of attention. He further argued that "the limited capacity and the arousal system must be closely related" (p. 10).

4.3 Fatigue

Generally, fatigue may be defined as the likelihood of falling asleep, but there is more to the condition than sleepiness alone. The fatigue state can be defined by its physical aspects, which are associated with muscular fatigue, a state caused by prolonged contraction of a muscle, a lack of oxygen, and an increased level of blood and muscle lactic acid (Craig and Tran 2012). However, for the purposes of this report, we are more concerned with the mental, behavioral, and physiological components associated with fatigue. A classic definition of fatigue in this context comes from Brown (1994), who described the fatigue state as a "subjectively experienced disinclination to continue performing the task at hand". Moreover, fatigue has also been closely linked to stress and is sometimes referred to as a stress-related condition (Bultmann et al. 2002). Therefore, we define fatigue as a neurophysiological state that occurs when a person is feeling tired or drowsy, or to the extent that they have a physical or cognitively reduced capacity to function, which may result in decreased performance, motivation, and generally negative emotions.

Fatigue, in part, reflects fundamental changes in neural function (Saper et al. 2005), but is also largely dependent upon the individual's level of subjective interest in the task (e.g., task engagement) as well as high-level cognitive processes relating to motivation regulation (Hockey 1997). Additionally, fatigue can also stem from different sources such as 1) an insufficiency or poor quality of sleep, 2) a byproduct of the 24-h circadian cycle in wakefulness and alertness, or 3) the task itself. A variety of task factors may influence the onset of fatigue (Ackerman et al. 2012). For example, high workload and monotonous tasks, which do not generally allow the individual to implement their own compensatory strategies, appear to be more susceptible to fatigue effects (Matthews et al. 2010a). Conversely, tasks that offer high levels of challenge and intrinsic interest can be highly fatigue resistant (Holding 1983; Neubauer et al. 2014).

4.3.1 Fatigue Measurement

Traditionally, fatigue research has been shaped by the tripartite division between physiological outcomes proposed by Bartley and Chute (1947), such as muscular fatigue, subjective feelings of tiredness and discomfort, and decrements in performance. The most obvious form of physiological fatigue is muscular fatigue resulting from prolonged physical exertion. Subjective fatigue may include an individual's awareness of bodily discomfort, but it also refers to the experience of mental states, including tiredness, sleepiness, apathy, and mind-wandering. Regarding objective measurement, the Psychomotor Vigilance Test (see Balkin et al. [2004]) provides a convenient, widely used measure of sleepiness based on reaction-time assessment and has been widely used in research on sleep and performance. Additionally, the performance decrements typically seen within vigilance research are also said to reflect an increased fatigue state.

4.3.2 Observable Correlates of Fatigue

Fatigue can be detected psychophysiologically (Wohleber et al. 2016) through EEG measures (Borghini et al. 2014), the PERCLOS (Wierwille et al. 1994), ECG measures (Borghini et al. 2014), and fluctuations in pupil diameter (Lüdtke et al. 1998). However, EEG measures have perhaps been researched the most extensively with regard to this particular state (Santamaria and Chiappa 1987; Wijesuriya et al. 2007). For example, within EEG research, the fatigue state has been found to be reliably associated with relative increases in slow wave activity (e.g., delta and theta waves) that are commonly experienced outwardly as an increase in drowsiness and sleepiness and are also inversely related to cortical arousal (Lal and Craig 2002; Craig and Tran 2012).

A systematic review of studies using a variety of experimental paradigms showed two consistent effects associated with reduced cortical arousal and hence fatigue. Power reliably increased in two frequency bands: theta (4–7.5 Hz) and alpha (8–13 Hz). Theta is linked to drowsiness and loss of alertness; alpha with a relaxed but wakeful state. Additionally, Dirnberger et al. (2004) studied the relationship between movement-related cortical potentials and fatigue, and found that lower amplitudes in movement potentials were associated with greater levels of fatigue. They suggested that this finding supports the notion that fatigue reduces cognitive capability and in turn cortical arousal. Lehmann et al. (1995) conducted canonical analyses between subjective cognitions and EEG spectral power profiles with four channels of EEG. They found significant pairs of variables that support the finding that fatigue is associated with reduced cortical arousal. For instance, they found 2–6-Hz (i.e., theta) activity to be significantly associated with cognition, reflecting reduced cognitive capacity (e.g., lacking orientation, low recall ability). Perhaps most importantly are the implications associated with these findings. For example, Gevins et al. (1990) argued that the changes associated with fatigue in the theta, alpha, and perhaps delta frequency bands could be seen prior to any deterioration in performance. Therefore, it appears that EEG may be used to provide an early warning for the onset of fatigue. That is, performance decrements are preceded in time by these increases in theta and alpha waves, suggesting brain wave activity may be a sensitive indicator of fatigue and its associated performance decrements.

Additionally, ECG studies have utilized decreased HRs as a measure of lower arousal, or fatigue (Borghini et al. 2014). More specifically, it has been argued that an increase in HRV may be indicative of a high fatigue state or higher emotion regulation (Wohlebler 2018). In fact, O'Hanlon (1972) found that HRV increased in response to time on task, an indicator of a fatigue state, and consequently manifested as a performance decrease and an apparent vigilance decrement, but then decreased when sudden events alerted drivers and gained their attention, resulting in an increase in their HR. Additionally, changes in HR reflect changes in autonomic arousal and sympathetic influences that govern emotional responses and decision-making. For example, given that task-induced fatigue and drowsiness are characterized as low-arousal states, a decline in average HR should be indicative of these changes (Borghini et al. 2014). In fact, studies within driving have shown significant declines in HR when drivers enter a fatigue state (Jap et al. 2009).

Finally, ocular metrics such as PERCLOS and changes in pupil diameter have also been studied in the context of fatigue state detection. As previously mentioned, PERCLOS is defined as the proportion of time that a person's eyes are more than 80% closed (Dinges et al. 1998). Researchers have classified a PERCLOS score of 15% or more as indicative of a person that is "drowsy", 7.5%–15% indicative of a "questionable state", and 7.5% or less indicative that the person is "awake" (Wierwille et al. 1994). The PERCLOS measure can serve as an initial measure of fatigue or drowsiness and reflects slow eyelid closures (<500 ms) rather than faster blinks, which are not computed for this measure (Wierwille et al. 1994; Kozak et al. 2005). In fact, Schleicher et al. (2008) define lid closures that are greater than 500 ms as "microsleeps". Finally, increases in the PUI have also been associated with sleepiness and fatigue (Lüdtke et al. 1998).

4.4 Vigilance

The term "vigilance" has been used in different ways by different groups of researchers. For example, neurophysiologists sometimes refer to the term vigilance as a measure of arousal within the sleep–wake spectrum, without the mention of any cognitive or behavioral response (Oken et al. 2006). Within this school of thought, changes in vigilance are primarily assumed to reflect activity in corticothalamic networks underlying the sleep–wake dimension (Steriade 2000). Conversely, within psychology and the cognitive science domain, the term has been used to describe an individual's ability to sustain attention toward a particular task over a period of time (Davies and Parasuraman 1982). For the purposes of this report, we adhere to the previous school of thought.

Vigilance, or "sustained attention", tasks require observers to maintain their focus of attention and detect the appearance of critical signals over prolonged periods of time (Warm et al. 2008b). Cognitively, sustained attention is most closely related to alertness. Alertness is also a term that overlaps with arousal and includes aspects of cognitive processing. In fact, some researchers have argued that arousal is an aspect of vigilance, and that the two constructs are very closely related, as seen in sleep-deprivation research (Parasuraman et al. 1998). More specifically, alertness can be measured via tonic (synonymous to vigilance and sustained attention [Posner and Peterson 1990]) and phasic (associated with orienting responses [Sokolov 1963]) changes. A key finding in vigilance research is that sustained attention is fragile and wanes over time. This is reflected in what is known as the vigilance decrement, a decline in the speed and accuracy of signal detections with time on task (Mackworth 1948; Warm et al. 2008b). Davies and Parasuraman (1982) reviewed the numerous factors that influence the vigilance decrement. These include task demand factors, such as memory load and stimulus event rate; variables that influence motivation, such as performance feedback; and adverse environmental conditions (Hancock 1984).

4.4.1 Vigilance Measurement

Typically, vigilance is measured via performance outcomes and is evident when individuals begin to make mistakes on a task, over time (i.e., the vigilance decrement). The resource model proposed by Davies and Parasuraman (1982) is a major conceptual framework for understanding the vigilance decrement. According to that view, the need to make continuous signal-to-noise discriminations depletes information-processing assets or reservoirs of energy that cannot be replenished in the time available, hence, the temporal decline in performance efficiency. Support for the resource model comes from findings that vigilance tasks impose a high level of perceived mental workload on observers, which increases with increments in psychophysical demand; that vigilance tasks promote high levels of stress; and that the decrement is correlated with observers' feelings of mental exhaustion (Warm et al. 2008b, 2015). Physiological and subjective reports also confirm that vigilance tasks reduce task engagement and increase distress and that these changes rise with increased task difficulty (Warm et al. 2008b).

4.4.2 Observable Correlates of Vigilance

It should be noted that observable correlates of vigilance have been hard to explicitly define, mostly due to the different schools of thought regarding this state, and because there are other factors that contribute to vigilance, rather than discrete changes in arousal. In fact, some researchers have argued that the term "tonic alertness" may be more fitting as it more clearly outlines physiological changes that occur during vigilance tasks (Oken et al. 2006). Furthermore, it appears that alertness and sustained attention have similar underlying brain processes, such as the thalamo-cortical pathways associated with the sleep–wake state. As such, most of the research assessing vigilance has relied on experimental paradigms that impact drowsiness and sleep-deprivation. However, the most commonly studied physiological modalities for vigilance measurement have focused on EEG (Stikic et al. 2011), various measures of eye movement (Unsworth and Robison 2016), and ANS activity (Steriade 2000).

During vigilance tasks, EEG signals have been found to be predictive of performance (Duta et al. 2010; Stikic et al. 2011). Overall, decreased vigilance has been associated with increased slow wave frequency and decreased amplitude of event-related potentials of the EEG. Conversely, during maximal attention, "awake" frequencies such as increased alpha wave activity have been recorded (Pfurtscheller and Aranibar 1977). Vigilance has also been associated with changes in ANS activity such as decreased HR, while diminished levels of vigilance are also indicated by increased HRV. Conversely, alerting events that rapidly increase vigilance are accompanied by sudden increases in HR and decreased HRV (O'Hanlon 1972). Electrodermal indicators of vigilance vary between labiles and stabiles, with labiles exhibiting relatively better performance in sustaining attention (Crider and Augenbraun 1975; Hastrup 1979; Davies and Parasuraman 1982; Vossel and Rossman 1984; Munro et al. 1987). Additionally, vigilance decrements are relatively more pronounced in the SCR and habituation rate of stabile's EDA (Munro et al. 1987; Koelega 1990) and also in combination with measures of eyelid closure and head nodding (St John et al. 2006). Moreover, Hopstaken et al. (2015) and Van Orden et al. (2000) both found gradual decreases in pupil diameter as well as perceptual sensitivity, and thus performance, during a sustained attention task. However, it should be noted that these approaches typically utilize regression of many features to predict vigilance, making concise description of the relation between vigilance and these sensed modalities challenging. Thus, while measures used in isolation have been found to predict changes in vigilance and subsequent task performance, several researchers have suggested that a combination of measures may be more sensitive to attentional states and, hence, produce better vigilance predictions (Van Orden et al. 2000).

4.5 Trust

It has long been believed that human trust perception is a primary determinant of human–autonomy interactions and further presumed that calibrating trust can lead to appropriate decisions regarding control authority. However, attempts to improve joint system decision-making by calibrating trust have not yet provided a generalizable solution. To address this, it is critical to first understand how we measure trust in human–autonomy teams and second use that measurement to identify when and how to calibrate team trust (for a full review of these measurement methods, see Krausman et al. [in press]). Preliminary research suggests that a multi-method approach that includes assessment of self-report, communication, behavior, and physiology together explains and expands on team performance ratings during joint human–autonomy team operations (Schaefer et al. 2019a; Milner et al. 2020).

4.5.1 Trust Measurement

A number of self-report trust scales, with varying assessment purposes, are available, though choosing self-report trust scales that address system trust and team trust is generally preferable in the context of human–autonomy teams. For this context, specifically in domain of Army lethality, scales include a system trust scale that focuses on system intelligence, safety, autonomy, trustworthiness, and use (Schaefer et al. 2012), and a team readiness questionnaire that addresses the team's readiness, confidence in specific types of autonomy, self-confidence, trust in the autonomy, and trust in the team (Schaefer et al. 2019a, 2019b; Milner et al. 2020). Other research on trust in autonomous vehicle driving aids (see Neubauer et al. [2000]) has supported using standard human–automation system trust scales such as the Checklist for Trust between People and Automation (Jian et al. 2000) or the System Trust Scale (Muir and Moray 1996). However, these scales are limited when it comes to intelligent systems. To address some of these limitations, modifications have been made these standard automation trust scales. For example, modifications to the Checklist for Trust between People and Automation have been made to address four functional areas of human–autonomy teaming: gathering or filtering information, integrating and displaying analyzed information, suggesting or making decisions, and executing actions (Wright et al. 2020). More recent scales, such as the Trust Perception Scale-Human-Robot Interaction (Schaefer 2016), address some of these limitations by expanding the item pool to address independent and interdependent teaming factors that extend across all robotic domains and joint team operations.

Overall, these scales explain self-reported trust-based perceptions, but can be strengthened through behavioral and communication analysis. The main difficulty with behavioral analysis is that it varies by task. Given that, some critical functions that may influence trust include attention management (e.g., eye movement, mean saccade amplitude, horizontal gaze deviation [He et al. 2011; Chen and Barnes 2012; Gold et al. 2015]), proximity (MacArthur et al. 2017; Schaefer et al. 2019c), or control authority (e.g., percent of time autonomy has control authority, per appropriate control modality [Spain and Bliss 2008]). Communication analysis can thus be used to further characterize the interaction and coordination patterns of human–autonomy teams that help explain team trust and cohesion (e.g., aggregate communication flow, relational event modeling, and language-similarity measures [Baker et al. 2020]).

4.5.2 Observable Correlates of Trust

Physiological measures, specifically EDA and HRV, have shown some promise as potential indicators that can be used to identify when changes in trust-based decision-making can occur. Although a one-to-one discrimination between EDA measures and specifics trust states cannot be directly asserted, several controlled experimental paradigms have linked changes in tonic and phasic EDA to changes in individual states such as stress and engagement within autonomous environments (Mower et al. 2007). In this context, increases in both tonic and phasic EDA have been shown to relate to increases in anxiety and cognitive effort (Shi et al. 2007; Zhao et al. 2015). Additionally, work by Bethel (2007) pioneered the use of EDA measurement during human–robot interaction studies. They found that tonic EDA measures increased along with increased engagement with the robot. Furthermore, Montague et al. (2014) employed a dyadic interaction trust paradigm and found that the higher the users' individual ratings of trust in technology were, the more their individual EDA measures agreed. In other words, if EDA levels of one subject were low, and both subjects were in a trust state, the second subject's EDA would be expected to be low as well. Moreover, in a pilot study, Sanders et al. (2012) found higher EDA and lower subjective trust ratings during interaction with an unreliable robot than with a reliable robot. These general findings are important within human–autonomy teams because such autonomous agents need to accurately assess changes in physiology so that they can respond appropriately. For example, if an autonomous system suddenly failed in some way (e.g., likely considered a negative, undesirable event), a subsequent sympathetic activation and an increase in tonic EDA should be present, and the degree or magnitude of the SCR would indicate the significance of the stimulus, in this case, the level of stress, frustration, or even anxiety associated with the negative event.

Additionally, HR and HRV measures are often used in conjunction with other sensing systems to infer the effect of a stimulus, cognitively and affectively, on a subject. Following a stimulus, an acute decrease in HRV, along with a simultaneous phasic response has been associated with orienting behavior (Figner and Murphy 2011), which may allow for an inference that an event was salient to the subject. HRV may also be used in conjunction with EDA to infer levels of workload and trust (Matthews et al. 2005; Mehler 2009; Montague et al. 2014). For example, in a state of high trust, it is unlikely that one would feel anxious, and therefore HRV should be high and HR and tonic EDA levels would be low. However, if there is an increase in cognitive workload and anxiety associated with the process of maintaining situation awareness such as might occur in a state of low trust, it is likely HRV would fall and NS-SCR would rise along with tonic EDA levels. However, this particular trend was found within a high-risk, intelligent teaming scenario and may not generalize outside this context. Finally, Mitkidis et al. (2015) found that in human–human interactions during a joint action task, HR arousal and HR synchrony between individuals was indicative of increased trust.

5. Conclusions and Limitations

This report outlines the importance of utilizing multimodal psychophysiological methods for human state detection. More specifically, one primary goal of this work focuses on the use of multimodal sensor fusion for the continuous assessment of human state in real time using noninvasive, wearable systems to support fieldable estimation of those states. This is an important area to focus on as systems are being developed to implement adaptive autonomous teammates to respond to changes in human state that indicate a precursor to undesirable decisions (e.g., unfavorable levels of stress, fatigue, workload, and/or trust). Additionally, we have presented other challenges associated with using subjective or even performance-based metrics, which may not capture the real-time dynamics associated with state change. Therefore, we seek to perform inference of these latent mental states from multimodal physiological and behavioral outputs (e.g., ECG, EDA, EEG, eye tracking, or facial expressions), specifically to improve interactions and performance in human–autonomy teams.

However, when discussing the use of psychophysiology as a means to estimate human state, it is critical to also present the caveats associated with these methodologies. Perhaps most importantly, it should not be assumed that an existing relationship between one or more elements of a psychological or physiological domain be held constant across all situations or individuals. In other words, a relationship between a theoretical construct and physiological and behavioral data may have a limited range of validity due to the fact that such relationships are only clear within well-controlled experimental paradigms and may actually be the result of other antecedent factors. In this context, it is vital to enact procedures that hold elements of the psychological domain constant in order to determine which of the observed changes in physiological response are likely attributable to relevant elements and which are believed to covary with irrelevant elements. As such, a further goal within this area of research is to use these procedures to account for or remove these irrelevant sources of variance to isolate the robust, if not invariant, relationship among psychological, physiological, and behavioral quantities of interest. Additionally, we have attempted to emphasize in the methods outlined previously that these relationships are simply a way to provide a more rigorous approach to understanding and modeling psychophysiological processes. However, researchers still need to be aware that these relationships are approximations and generalizations that can vary across individuals, time, and contexts. Still, even coarse information about the connection between observable and unobservable human characteristics can be leveraged to improve prediction of their behavior.

Overall, within this report, we would like to stress that there does appear to be, at a minimum, a transactional relationship between the individual and environment, which unfolds as changes in psychological state are exhibited via physiological and behavioral manifestations. Although a direct, invariant psychophysiological relationship provides the best generality, it is our contention that physiological and behavioral markers, covariates, and observable outcomes also provide important information about specific unobservable variables of interest within the psychological domain. Therefore, this report serves as a reference that identifies current psychophysiological relationships and their associated inferences that are applicable to human–autonomy teams, via noninvasive and wearable sensor technologies. We have also attempted to visually and succinctly summarize these directional relationships in Appendix B, Table 10 of this document. Finally, this report has addressed various questions relating to selecting the appropriate measurements for variables of interest, the importance of addressing contextual and individual variability on psychophysiological relationships, and how measures of physiological and behavioral signals can be utilized to index various psychological factors such as stress, fatigue, workload, vigilance, and trust.

Finally, this report has argued for the use of multimodal sensing and modeling techniques within psychophysiological research to more effectively predict state-based interactions within the context of human–autonomy teams. Integration of eye-tracking, EDA, EEG, and/or ECG measures allows researchers to more holistically and comprehensively infer internal human states that impact those interactions than is possible with any one measure individually. Our primary aim has been to support the design of psychophysiological models, informed by a priori knowledge, by characterizing the robust relationships between those states and their physiological and behavioral indicators to enable more powerful methods for human-centered sensor fusion and ultimately enable effective and resilient human–autonomy teaming.

6. References

Ackerman PL, Calderwood C, Conklin EM. Task characteristics and fatigue. In: Matthews G, Desmond PA, Neubauer C, Hancock PA (eds). Handbook of operator fatigue. Aldershot (UK): Ashgate Press; 2012. pp. 139–154, 2010.

Allgöwer F, Zheng A. (eds.). Nonlinear model predictive control (Vol. 26). Birkhäuser; 2012.

Baker AL, Fitzhugh SM, Forster D, Scharine A, Brewer RW, Krausman A, Schaefer KE. Team trust in human–autonomy teams: analysis of crew communication during manned-unmanned gunnery operations. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2020. Report No.: ARL-TR-8969.

Balkin TJ, Bliese PD, Belenky G, et al. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. J Sleep Res. 2004;13:219–27.

Baltrušaitis T, Robinson P, Morency LP. OpenFace: an open source facial behavior analysis toolkit. In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV); 2016 Mar. pp. 1–10. IEEE.

Bartlett B, Estivill-Castro V, Seymon, S. Dogs or robots: why do children see them as robotic pets rather than canine machines?. In Proceedings of the fifth conference on Australasian user interface. 2004 Jan;28:7–14.

Bartley SH, Chute E. Fatigue and impairment in man. New York (NY): McGraw Hill; 1947.

Batrinca L, Stratou G, Shapiro A, Morency LP, Scherer S. Cicero-towards a multimodal virtual audience platform for public speaking training. In International workshop on intelligent virtual agents. Berlin, Heidelberg (Germany): Springer; 2013 Aug. pp. 116–128.

Beatty J. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin. 1982;91:276–292.

Beer Jenay M, Fiske Arthur D, Rogers Wendy A. Toward a framework for levels of robot autonomy in human-robot interaction. Journal of Human-Robot Interaction. 2014;3(2):74–99.

Ben-Shakhar, G. Standardization within individuals: a simple method to neutralize individual differences in skin conductance. Psychophysiology. 1985;22(3):292–299.

Berntson GG, Cacioppo JT, Binkley PF, Quigley KS, Uchino BN, Fieldstone A. Autonomic cardiac control. Psychological stress and cardiac response in autonomic space as revealed by pharmacological blockades. Psychophysiology. 1994;31:599–608.

Berntson GG, Cacioppo JT, Quigley KS. The metrics of cardiac chronotropism: biometric perspectives. Psychophysiology. 1995;32(2):162–171.

Berntson GG, Cacioppo JT, Quigley KS. Cardiac psychophysiology and autonomic space in humans: empirical perspectives and conceptual implications. Psychological Bulletin. 1993;114(2):296.

Berntson GG, Thomas Bigger Jr J, Eckberg DL, Grossman P, Kaufmann PG, Malik M, Van Der Molen MW. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997;34(6):623–648.

Berntson GG, Quigley KS, Lozano D. Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Eds.). Handbook of psychophysiology. Cambridge University Press; 2007. p. 182–210.

Bethel JL, Murphy RR, Salomon K. Psychophysiological experimental design for use in human-robot interaction studies. In: International Symposium on Collaborative Technologies and Systems; 2007.

Biros DP, Daly M, Gunsch G. The influence of task load and automation trust on deception detection. Group Decision and Negotiation. 2004;13:173–189.

Bonato P. Wearable sensors and systems. IEEE Engineering in Medicine and Biology Magazine. 2010;29(3):25–36.

Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev. 2014;44:58–7.

Boucsein W, Backs RW. Engineering psychophysiology as a discipline: Historical and theoretical aspects. Engineering Psychophysiology. Issues and Applications. 2000;3–30.

Bradford JC, Lukos JR, Passaro A, Ries A, Ferris DP. Effect of locomotor demands on cognitive processing. Scientific Reports. 2019;9(1):1–12.

Bradford JC, Lukos JR, Ferris DP. Electrocortical activity distinguishes between uphill and level walking in humans. Journal of Neurophysiology. 2016;115(2):958–966.

Bradley MM, Miccoli L, Escrig MA, Lang PJ. The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology. 2008;45:602–607.

Brown I. Driver fatigue. Hum Factors. 1994;36:298–314.

Bultmann U, Kant IJ, Schroer CAP, Kasl SV. The relationship between psychosocial work characteristics and fatigue and psychological distress. International Archives of Occupational and Environmental Health. 2002;75:259–266.

Burke CS, Sims DE, Lazzara EH, Salas E. Trust in leadership: A multi-level review and integration. The Leadership Quarterly. 2007;18(6):606–632.

Byrne EA, Parasuraman R. Psychophysiology and adaptive automation. Biological Psychology. 1996;42:249–268.

Cacioppo JT, Berntson GG. Relationship between attitudes and evaluative space: a critical review, with emphasis on the separability of positive and negative substrates. Psychological Bulletin. 1994;115(3):401.

Cacioppo JT, Tassinary LG, Berntson, G., editors. Handbook of psychophysiology. Cambridge University Press; 2007.

Cacioppo JT, Tassinary LG. Inferring psychological significance from physiological signals. American Psychologist. 1990;45(1):16.

Cacioppo JT, Tassinary LG, Berntson GG. Psychophysiological science. Handbook of Psychophysiology. 2000;2:3–23.

Camacho EF, Alba CB. Model predictive control. Springer Science and Business Media; 2013.

Chen JYC, Barnes MJ. Supervisory control of multiple robots: Effects of imperfect automation and individual differences. Human Factors. 2012 Apr;54:157–174.

Chen X, Duan Y, Houthooft R, Schulman J, Sutskever I, Abbeel P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems; 2016. p. 2172–2180.

Chollet M, Wörtwein T, Schauerte B, Morency LP, Stiefelhagen R, Scherer S. Multimodal public speaking performance assessment. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction; 2015 Nov. p. 43–50.

Cohen Hoffing RAC, Lauharatanahirun N, Forster DE, Garcia JO, Vettel JM, Thurman SM. Dissociable mappings of tonic and phasic pupillary features onto cognitive processes involved in mental arithmetic. PloS ONE. 2020;15(3):e0230517.

Cosenzo KA, Parasuraman R, Novak A, Barnes M. Implementation of automation for control of robotic systems. Aberdeen Proving Grounds (MD): Army Research Laboratory (US); 2006. Report No.: ARL-TR-3808.

Craig A, Tran Y. The influence of fatigue on brain activity. In: Matthews G, Desmond PA, Neubauer C, Hancock PA., editors. The handbook of operator fatigue. Burlington (NJ): Ashgate Publishing Company; 2012. p. 185–196.

Crider A, Adgenbraun CB. Auditory vigilance correlates of electrodermal response habituation speed. Psychophysiology. 1975;12(1):36–40.

Cuthbert BN, Bradley MM, Lang PJ. Picture media and emotion: effects of a sustained affective context. Psychophysiology. 1996;33(6):662–670.

Dawson ME, Schell AM, Filion DL. The electrodermal system. In: Systemic psychophysiology. Cambridge University Press; 2017.

Davies DR, Parasuraman R. The psychology of vigilance. Academic Press; 1982.

Delliaux S, Delaforge A, Dehara J-C, Chaumet G. Mental workload alters heart rate variability, lowering non-linear dynamics. Frontiers in Physiology. doi.org/10.3389/fphys.2019.00565.

Denison RN, Parker JA, Carrasco M. Modeling pupil responses to rapid sequential events. Behav Res. 2020. https://doi.org/10.3758/s13428-020-01368-6.

Desmond PA, Hancock PA. Active and passive fatigue states. In: Hancock PA, Desmond PA, editors. Stress, workload, and fatigue. Mahwah (NJ): Lawrence Erlbaum Associates; 2001. p. 455–465.

DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, Lucas G. SimSensei Kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems; 2014 May. p. 1061–1068.

Dinges DF, Mallis MM, Maislin G, Powell JW. Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management (No. DOT-HS-808-762). Washington (DC): National Highway Traffic Safety Administration (US); 1998.

Dirnberger G, Duregger C, Lindinger G, Lang W. Habituation in a simple repetitive motor task: a study with movement-related cortical potentials. Clin Neurophysiol. 2004;115:378–384.

Doesburg SM, Green JJ, McDonald JJ, Ward LM. Rhythms of consciousness: binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. PloS ONE. 2009;4(7):e6142.

Dong W, Moan T, Gao, Z. Long-term fatigue analysis of multi-planar tubular joints for jacket-type offshore wind turbine in time domain. Engineering Structures. 2011;33(6):2002–2014.

Donmez B, Boyle L, Lee JD, McGehee D. Drivers' attitudes towards imperfect distraction mitigation strategies. Trans Res F: Psychol Behav. 2006;9(6):387–398.

Duchowski AT, Krejtz K, Krejtz I, Biele C, Niedzielska A, Kiefer P, Giannopoulos I. The index of pupillary activity: measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018 Apr. p. 1–13.

Dukkipati RV. Analysis and design of control systems using MATLAB. New Age International; 2006.

Duta M, Alford C, Wilson S, Tarassenko L. Neural network analysis of the mastoid EEG for the assessment of vigilance. International J Human-Computer Interaction. 2004;17(2):171–195. doi:10.1207/s15327590ijhc1702_4.

Eckberg DL, Berntson GG, Thomas Bigger Jr J, Grossman P, Kaufmann PG, Malik M, Van Der Molen MW. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology. 1997;34(6):623–648.

Eckberg DL, Taylor JA, Carr DL, Myers CW. Mechanisms underlying very-low-frequency RR-interval oscillations in humans. Circulation. 1998;98(6):547–555.

Ekman P, Friesen W. Facial action coding system: a technique for the measurement of facial movement. Palo Alto (CA): Consulting Psychologists Press; 1978.

Fairclough SH, Venables L. Prediction of subjective states from psychophysiology: a multivariate approach. Biological Psychology. 2006;71(1):100–110.

Figner B, Murphy RO. Using skin conductance in judgment and decision making research. a handbook of process tracing methods for decision research. In: Schulte-Mecklenbeck M, Kuehberger A, Ranyard R, editors. A handbook of process tracing methods for decision research. New York (NY): Psychology Press; 2011.

Foxe JJ, Snyder AC. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. Frontiers in Psychology. 2011;2:154.

Gabay S, Pertzov Y, Henik A. Orienting of attention, pupil size, and the norepinephrine system. Atten Percept Psychophys. 2011;73:123–129.

Gaetz W, Edgar JC, Wang DJ, Roberts TP. Relating MEG measured motor cortical oscillations to resting γ-aminobutyric acid (GABA) concentration. Neuroimage. 2011;55(2):616–621.

Gevins AS, Bressler SL, Cutillo BA, Illes J, Miller JC, Stern J, Jex HR. Effects of prolonged mental work on functional brain topography. Electroencephalogr. Clin Neurophysiol. 1990;76(4):339–350.

Gildöf K, Holmberg N, Sandberg H. The use of eye-tracking and retrospective interviews to study teenagers' exposure to online advertising. Visual Communication. 2012;11:329–345. doi:10.1027/1016-9040.

Glaholt MG, Reingold EM. Eye movement monitoring as a process tracing methodology in decision making research. Journal of Neuroscience, Psychology, and Economics. 2011;4.

Golan Z, Neufeld MY. Individual dierences in alpha rhythm as characterizing temperament related to cognitive performances. Personality and Individual Differences. 1996;21(5):775–784.

Goldman RI, Stern JM, Engel J, Jr, Cohen MS. Simultaneous EEG and fMRI of the alpha rhythm. NeuroReport. 2002;13:2487–2492.

Gold C, Körber M, Hohenberger C, Lechner D, Bengler K. Trust in automation – Before and after the experience of take-over scenarios in a highly automated vehicle. Procedia Manufacturing. 2015;3:3025–3032.

Goodrich M, Olsen D. Seven principles of efficient human robot interaction. Proc IEEE International Conference on Systems, Man and Cybernetics. 2003:3943–3948.

Granholm E, Asarnow RF, Sarkin AJ, Dykes KL. Pupillary responses index cognitive resource limitations. Psychophysiology. 1996;33(4):457–461.

Gremillion GM, Donavanik D, Neubauer CE, Brody JD, Schaefer K.E. Estimating human state from simulated assisted driving with stochastic filtering techniques. In: International Conference on Applied Human Factors and Ergonomics; 2018 July. p. 113–125. Springer.

Grings WW, Schell AM. Magnitude of electrodermal response to a standard stimulus as a function of intensity and proximity of a prior stimulus. Journal of Comparative and Physiological Psychology. 1969;67(1):77–82. https://doi.org/10.1037/h0026651.

Grossman P, Karemaker J, Wieling W. Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: the need for respiratory control. Psychophysiology. 1991;28(2):201–216.

Grossman P, Kollali M. Respiratory sinus arrhythmia, cardiac vagal tone, and respiration: within-and between-individual relations. Psychophysiology. 1993;30(5):486-495.

Gu H, Ji Q. An automated face reader for fatigue detection. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition; 2004 May. p. 111–116. IEEE.

Güntekin B, Başar E. Event-related beta oscillations are affected by emotional eliciting stimuli. Neuroscience Letters. 2010;483(3):173–178.

Hancock PA. Environmental Stressor. In: Warm JS (eds.). Sustained attention in human performance. New York (NY): Wiley; 1984. p. 103–142.

Hancock PA, Warm JS. A dynamic model of stress and sustained attention. Hum Factors. 1989 Oct;31(5):519-37. doi: 10.1177/001872088903100503. PMID: 2625347.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human mental workload. Amsterdam (Netherlands): North-Holland Press; 1988. p. 239–250.

Hastrup JL. Effects of electrodermal lability and introversion on vigilance decrement. Psychophysiology. 1979;16(3):302–310.

Hazlett H, Dawson ME, Schell AM, Nuechterlein KH. Electrodermal Activity as a Schizophrenia. Biol Psychiatry. 1997;997(41):111–113.

He J, Becic E, Lee, Y-C, McCarley JS. Mind wandering behind the wheel: Performance and oculomotor correlates. Human Factors. 2011;53:13–21

Heathers JA. Everything Hertz: methodological issues in short-term frequency-domain HRV. Frontiers in Physiology. 2014;5:177.

Hess EH, Polt JM. Pupil size in relation to mental activity during simple problem-solving. Science. 1964 Mar 13;143(3611):1190-2. doi: 10.1126/science.143.3611.1190. PMID: 17833905.

Higgins I, Matthey L, Pal A, Burgess C, Glorot X, Botvinick M, Mohamed S, Lerchner A. beta-VAE: learning basic visual concepts with a constrained variational framework. ICLR; 2017.

Hjortskov N, Rissén D, Blangsted AK, Fallentin N, Lundberg U, Søgaard K. The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol. 2004 June;92(1-2):84–89. doi: 10.1007/s00421-004-1055-z. Epub 2004 Feb 27. PMID: 14991326.

Hockey GJ. Compensatory control in the regulation of human performance under stress and high workload: a cognitive–energetical framework. Biological Psychology. 1997;45:73–93.

Holding DH. Fatigue. In G.R.J. Hockey, editor. Stress and fatigue in human performance. Chichester (UK): John Wiley; 1983.

Hopstaken JF, van der Linden D, Bakker AB, Kompier MA. The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. Biol Psychol. 2015;110:100–106.

Hughes AM, Hancock GM, Marlow SL, Stowers K, Salas E. Cardiac measures of cognitive workload: a meta-analysis. Human Factors. 2019;61(3):393–414. https://doi.org/10.1177/0018720819830553.

Jackson AF, Bolger DJ. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. Psychophysiology. 2014;51(11):1061–1071.

Jap BT, Lal S, Fischer P, Bekiaris E. Using EEG spectral components to assess algorithms for detecting fatigue. Expert Syst Appl. 2009;36:2352–2359.

Jepma M, Nieuwenhuis S. Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. Journal of Cognitive Neuroscience. 2011;23(7):1587–1596.

Jerath R, Crawford MW, Barnes VA. A unified 3D default space consciousness model combining neurological and physiological processes that underlie conscious experience. Frontiers in Psychology. 2015;6:1204.

Jennings JR. Bodily changes during attending. Psychophysiology: Systems, Processes, and Applications. 1986;268–289.

Jian JY, Bisantz AM, Drury CG, Llinas J. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics. 2000;4:53–71.

Joshi S, Li Y, Kalwani RM, Gold JI. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron. 2016;89(1):221–234.

Just MA, Carpenter PA. Eye fixations and cognitive processes. Cognitive Psychology. 1976;8.

Kahneman D. Attention and effort. Englewood Cliffs (NJ): Prentice Hall Inc.; 1973.

Kahneman D, Beatty J. Pupil diameter and load on memory. Science. 1966;154:1583–1585.

Kappenman ES, Luck SJ. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology. 2010;47(5):888–904.

Katkin ES. Electrodermal lability: A psychophysiological analysis of individual differences in response to stress. Stress and Anxiety. 1975;2:141–176.

Keesman KJ. System identification: an introduction. Springer Science and Business Media; 2011.

Kelly SP, Lalor EC, Reilly RB, Foxe JJ. Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. Journal of Neurophysiology. 2006;95(6):3844–3851.

Kerick SE, Oie KS, McDowell K. Assessment of EEG signal quality in motion environments. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2009. Report No.: ARL-TR-4866.

Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews. 1999;29:169–195.

Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences. 2012;16(12):606–617.

Koelega HS, Brinkman JA, Zwep B, Verbaten MN. Dynamic vs static stimuli in their effect on visual vigilance performance. Perceptual and Motor Skills. 1990;70(3):823–831.

Körber M, Cingel A, Zimmermann M, Bengler K. Vigilance decrement and passive fatigue caused by monotony in automated driving. Procedia Manufact. 2015;3:2403–2409.

Korn CW, Bach DR. A solid frame for the window on cognition: Modeling event-related pupil responses. Journal of Vision. 2016;16(3):28.

Kozak K, Curry R, Greenberg J, Artz B, Blommer M, Cathey L. Leading indicators of drowsiness in simulated driving. Proc Hum Factors Ergon Soc Annu Meet. 2005;49:1917–1921. SAGE Publications.

Kramer HC, Trejo LJ, Humphrey DG. Psychophysiological measures of workload: potential applications to adaptively automated systems. Mahwah (NJ): Lawrence Erlbaum Associates; 1996.

Krausman A, Baker A, Neubauer C, Lakhmani S, Gremillion G, Fitzhugh S, Wright J, Scharine A, Forster D, Metcalfe J, Schaefer, K. Team trust metrics for human–autonomy teams. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); in press.

Ladouce S, Donaldson DI, Dudchenko PA, Ietswaart M. Mobile EEG identifies the re-allocation of attention during real-world activity. Scientific Reports. 2019;9(1):1–10.

Lal SKL, Craig A. Driver fatigue electroencephalography and psychological assessment. Psychophysiology. 2002 May;39:1–9.

Landis EM. Micro-injection studies of capillary blood pressure in human skin. Heart. 1930;15:209–228.

Lang PJ, Greenwald MK, Bradley MM, Hamm AO. (). Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology. 1993;30(3):261–273.

Lazarus RS. Stress and emotion: a new synthesis. New York (NY): Springer; 1999.

Lehmann D, Grass P, Meier B. Spontaneous conscious covert cognition states and brain electric spectral states in canonical correlations. International Journal of Psychophysiology. 1995;19:41–52.

Loewenfeld IE, Lowenstein O. The pupil: anatomy, physiology, and clinical applications (vol. 2). Wiley-Blackwell; 1993.

Lüdtke H, Wilhelm B, Adler M, Schaeffel F, Wilhelm H. Mathematical procedures in data recording and processing of pupillary fatigue waves. Vision Research. 1998;38(19):2889–2896.

MacArthur KR, Stowers K, Hancock PA. Human-robot interaction: proximity and speed—slowly back away from the robot!. In Advances in human factors in robots and unmanned systems. Springer; 2017. p. 365–374.

Mackworth NH. The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology. 1948;1:6–21

Marathe A, Brewer R, Kellihan B, Schaefer KE. Leveraging wearable technologies to improve test and evaluation of human-agent teams. Theoretical Issues in Ergonomics Science. 2020;1–21.

Marshall SP. Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. US Patent 6,090,051. Washington DC: US Patent and Trademark Office, 2000.

Marshall SP. Identifying cognitive state from eye metrics. Aviat Space Environ Med. 2007;78:B165–B175.

Mascord DJ, Heath RA. Behavioral and physiological indices of fatigue in a visual tracking task. J Saf Res. 1992;23(1):19–25.

Matthews G, Davies DR, Stammers RB, Westerman SJ. Human performance: cognition, stress, and individual differences. Psychology Press; 2000a.

Matthews G, Desmond PA. Task-induced fatigue states and simulated driving performance. Q J Exp Psyc Hol Sect A. 2002;55(2):659–686.

Matthews G, Desmond PA. Stress and driving performance: implications for design and training. 2001.

Matthews G, Desmond PA, Neubauer C, Hancock PA. (eds.). The handbook of operator fatigue. Farnham, Surrey (UK): Ashgate Publishing; 2012a. ISBN: 978-0-7546-7537-2.

Matthews R, McDonald NJ, Trejo LJ. Psycho-physiological sensor techniques: an overview. In: 11th International Conference on Human Computer Interaction (HCII); 2005.

Matthews G, Szalma JL, Panganiban AR, Neubauer C, Warm JS. Profiling task stress with the Dundee Stress State Questionnaire. In: Cavalcanti L, Azevedo S, editors. Psychology of stress: new research. Hauppauge (NY): Nova Science Publishers; 2012b.

Matthews G, Warm JS, Reinerman LE, Langheim LK, Saxby DJ. Task engagement, attention and executive control. In: Gruszka A, Matthews G, Szymura B, editors. Handbook of individual differences in cognition:

attention, memory and executive control. New York (NY): Springer; 2010a. p. 205–230.

Matthews G, Warm JS, Reinerman LE, Langheim L, Washburn DA, Tripp L. Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. Journal of Experimental Psychology: Applied. 2010b;16:187–203.

Matthews G, De Winter J, Hancock PA. What do subjective workload scales really measure? operational and representation solutions to the divergence of workload measures. Theoretical Issues in Ergonomics Science; 2020b.

McAllister DJ. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of Management Journal. 1995;38(1):24–59.

McDowell K, Lin CT, Oie KS, Jung TP, Gordon S, Whitaker KW, Hairston WD. Real-world neuroimaging technologies. IEEE Access. 2013;1:131–149.

Mehler B. Benefiting from automation: trust and senior mobility. AgeLab MIT; 2009.

Melloni L, Molina C, Pena M, Torres D, Singer W, Rodriguez E. Synchronization of neural activity across cortical areas correlates with conscious perception. Journal of neuroscience. 2007;27(11):2858–2865.

Metcalfe, JS, AR Marathe, B Haynes, VJ Paul, GM Gremillion, K Drnec, C Atwater, et al. Building a Framework to Manage Trust in Automation. Micro- and Nanotechnology Sensors, Systems, and Applications IX; 2017; 10194:101941U. International Society for Optics and Photonics.

Milner A, et al. Identifying New Team Trust and Team Cohesion Metrics that Support Future Human–autonomy teams. In: Cassenti D, Scataglini S, Rajulu S, Wright J, editors. Advances in simulation and digital human modeling. AHFE 2020. Advances in intelligent systems and computing, vol 1206. Springer; 2021.

Mitkidis P, McGraw JJ, Roepstorff A, Wallot S. Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. Physiology and Behavior. 2015;149:101–106.

Montague E, Xu J, Chiou E. 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. 2014;44.

Morrison JG, Gluckman JP. Definitions and prospective guidelines for the application of automation. In: Mouloua M, Parasuraman R (eds), Human performance in automated systems: current research and trends. Hillsdale (NJ): Lawrence Erlbaum Associates; 1994. p. 256–263.

Mower E, Feil-Seifer DJ, Mataric MJ, Narayanan S. Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements. In: RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication; 2007 Aug. p. 1125–1130. IEEE.

Muir B, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in automated systems. Ergonomics. 1996;39:429–460.

Mulder G, Mulder LJM. Information processing and cardiovascular control. Psychophysiology. 1981;18:391–402.

Munro LL, Dawson ME, Schell AM, Sakai LM. Electrodermal lability and rapid vigilance decrement in a degraded stimulus continuous performance task. Journal of Psychophysiology. 1987.

Nagai Y, Goldstein LH, Critchley HD, Fenwick PBC. Influence of sympathetic autonomic arousal on cortical arousal: implications for a therapeutic behavioural intervention in epilepsy. Epilepsy Res. 2004;58:185–193.

Nelles O. Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer Science and Business Media; 2013.

Neubauer C, Gremillion G, Perelman B, La Fleur C, Metcalfe J, Schaefer-Lay K. How analysis of facial expressions explain affective state and trust-based decisions during interaction with autonomy aids. In: Proceedings of the 3rd International Conference on Intelligent Human Systems Integration: Integrating People and Intelligent Systems; 2020; Modena, Italy.

Neubauer C, Matthews G, Langheim L, Saxby D. Fatigue and voluntary utilization of automation in simulated driving. Hum Factors. 2012;54(5):734–746.

Neubauer C, Matthews G, Saxby D. Fatigue in the automated vehicle: do games and conversation energize or distract the driver? Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2014;58(1):2053–2057.

Neubauer C, Chollet M, Mozgai S, Dennison M, Khooshabeh P, Scherer S. The relationship between task-induced stress, vocal changes, and physiological state during a dyadic team task. Proceedings of the 19th ACM International

Conference on Multimodal Interaction; 2017a Nov; New York, NY. p. 426–432.

Neubauer C, Mozgai S, Chuang B, Woolley J, Scherer S. Manual and automatic measures confirm—intranasal oxytocin increases facial expressivity. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII); 2017b. p. 229–235. IEEE.

Nuechterlein KH, Dawson ME. A heuristic vulnerability/stress model of schizophrenic episodes. Schizophrenia Bulletin. 1984;10(2):300.

Nunan D, Sandercock GR, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. Pacing Clin Electrophysiol. 2010;33:1407–1417.

O'Hanlon JF. Heart rate variability: a new index of driver alertness/fatigue (no. 720141). Warrendale (PA): SAE; 1972.

Oken BS, Salinsky MC, Elsas SM. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin Neurophysiol. 2006;117(9):1885–1901. doi:10.1016/j.clinph.2006.01.017.

Oliveira AS, Schlink BR, Hairston WD, König P, Ferris DP. Proposing metrics for benchmarking novel EEG technologies towards real-world measurements. Frontiers in Human Neuroscience. 2016;10:188.

Parasuraman R, Bahri T, Deaton J, Morrison J, Barnes M. Theory and design of adaptive automation in aviation systems. Warminster (PA): Naval Air Warfare Center; 1992. Report No.: NAWCADWAR-92033-60.

Parasuraman R, Greenwood PM. Selective attention in aging and dementia. In: Parasuraman R (ed.). The attentive brain. The MIT Press; 1998. p. 461–487.

Parasuraman R, Wilson G. Putting the brain to work: neuroergonomics past, present, and future. Human Factors. 2008;50:468–474.

Parra F, Miljkovitch R, Persiaux G, Morales M, Scherer S. The multimodal assessment of adult attachment security: developing the biometric attachment test. Journal of Medical Internet Research. 2017;19(4):e100.

Patel M, Lal SK, Kavanagh D, Rossiter P. Applying neural network analysis on heart rate variability data to assess driver fatigue. Expert Syst Appl. 2011;38:7235–7242.

Pfurtscheller G, Aranibar A. Event-related cortical desynchronization detected by power measurements of scalp EEG. Electroencephalogr Clin Neurophysiol.

1977 June;42(6):817–826. doi: 10.1016/0013-4694(77)90235-8. PMID: 67933.

Phillips E, Schuster D, Ososky S, Jentsch F, Lebiere C, Evans WA. A research approach to shared mental models and situation assessment in future robot teams. Proc Human Factors and Ergonomics Society Annual Meeting. 2011 Sep;55(1):456–460. Los Angeles (CA): SAGE Publications.

Poole A, Ball LJ. Eye tracking in HCI and usability research. Encyclopedia of Human Computer Interaction. 2006;1:211–219.

Posner, M. I. and Petersen, S. E. The attention system of the human brain. Annual Review of Neuroscience. 1990;13: 25–42.

Prinzel LJ, Freeman FG, Scerbo MW, Mikulka PJ, Pope AT. Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. Hum Factors. 2003;45(4):601–614.

Quintana DS, Heathers JA. Considerations in the assessment of heart rate variability in biobehavioral research. Frontiers in Psychology. 2014;5:805.

Rozado D, Dunser A. Combining EEG with pupillometry to improve cognitive workload detection. Computer. 2015;48.10:18–25.

Sahayadhas A, Sundaraj K, Murugappan M. Drowsiness detection during different times of day using multiple features. Australas Phys Eng Sci Med. 2013;36:243–250.

Sanders TL, Schaefer KE, Yordon RE, Billings DR, Hancock PA. Classification of robot form: Factors predicting perceived trustworthiness. Proc Human Factors and Ergonomics Society Annual Meeting. 2012 Sep;56(1):1548–1552. Los Angeles (CA): SAGE Publications.

Santamaria J, Chiappa KH. The EEG of drowsiness in normal adults. Journal of Clinical Neuruphysiology. 1987;4:327–382.

Saper CB, Scammell TE, Lu J. Hypothalamic regulation of sleep and circadian rhythms. Nature. 2005;437:1257–1263.

Saxby DJ, Matthews G, Hitchcock EM, Warm JS, Funke GJ, Gantzer T. Effect of active and passive fatigue on performance using a driving simulator. Proc Hum Factors Ergon Soc Annu Meet. 2008;52:1751–1755.

Scerbo M. Theoretical perspectives on adaptive automation. In: Parasuraman R, Mouloua M (eds.). Automation and human performance: theory and applications. Mahwah (NJ): Erlbaum; 1996. p. 37–63.

Schaefer KE. Measuring trust in human robot interactions: development of the "trust perception scale – HRI." In: Lawless W, Mittu R, Wagner A, Sofge D. (eds.), The intersection of robust intelligence (RI) and trust in autonomous systems. Springer; 2016. p. 191–218.

Schaefer KE, Sanders TL, Yordon RE, Billings DR, Hancock PA. Classification of robot form: factors predicting perceived trustworthiness. In: Proceedings of the 56th Annual Human Factors and Ergonomics Society; 2012 Sep; Boston, MA. p. 1548–1552.

Schaefer KE, Brewer RW, Baker AL, Pursel ER, Gipson B, Ratka S, Giacchi J, Cerame E, Pirozzo K. US Army Wingman Joint Capability Technology Demonstration (JCTD): initial Soldier and Marine feedback on manned-unmanned gunnery operations. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2019b Mar. Report No.: ARL-TR-8663.

Schaefer KE, Baker AL, Brewer RW, Patton D, Canady J, Metcalfe JS. Assessing multi-agent human–autonomy teams: US Army Robotic Wingman gunnery operations. In: Proceedings of the SPIE Defense + Commercial Sensing, Micro- and Nanotechnology Sensors, Systems, and Applications XI Conference; 2019a May; Baltimore, MD. doi://10.1117/12.2519302.

Schaefer KE, Hill SG, Jentsch FG. Trust in human–autonomy teaming: a review of trust research from the US Army Research Laboratory Robotics Collaborative Technology Alliance. In: Chen J., editor. Advances in human factors in robots and unmanned systems. AHFE 2018. Advances in Intelligent Systems and Computing, 2019c;784. Springer.

Schleicher R, Galley N, Briest S, Galley L. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics. 2008;51:982–1010.

Schnall PL, Landbergis, PA, Baker D. Job strain and cardiovascular disease. Annu Rev Public Health. 1994;15:318–411.

Schnall PL, Pieper C, Schwartz JE, et al. The relationship between 'job strain,' workplace diastolic blood pressure, and left ventricular mass index: results of a case-control study. JAMA. 1990;263(14):1929–1935. doi:10.1001/jama.1990.03440140055031.

Scholtz J, Antonishek B, Young J. Evaluation of operator interventions in autonomous off-road driving. In: Proceedings of the NIST Performance Metrics for Intelligent Systems Workshop; 2003.

Scribner DR, Animashaun A, Culbertson W. Soldier performance in a moving command vehicle under manned, teleoperated, and teleoperated cruise control modes under day and night conditions. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2013. Report No.: ARL-TR-6450.

Scribner DR, Dahn DA. Comparison of soldier performance in a moving command vehicle under manned, teleoperated, and semi-autonomous robotic mine detector system control modes. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2008. Report No.: ARL-TR-4609.

Segerstrom SC, Nes LS. Heart rate variability reflects self-regulatory strength, effort, and fatigue. Psychol Sci. 2007;18(3):275–281.

Simon D. Optimal state estimation: Kalman, H infinity, and nonlinear approaches. John Wiley and Sons; 2006.

Shi Y, Ruiz N, Taib R, Choi E, Chen F. Galvanic skin response (GSR) as an index of cognitive load. In CHI’07 Extended Abstracts on Human Factors in Computing Systems; 2007.

Sokolov EN. Higher nervous functions: The orienting reflex. Annu Rcv Physio. 1963;25:545.

Spain RD, Bliss JP. The effect of sonification display pulse rate and reliability of operator trust and perceived workload during a simulated patient monitoring task. Ergonomics. 2008;51:1320–1337.

St John M, Risser Matthew R, Kobus David A. Toward a usable closed-loop attention management system: predicting vigilance from minimal contact head, eye, and EEG measures. Foundations of Augmented Cognition. 2006;12–18.

Steinfeld A, Fong T, Kaber D, Lewis M, Scholtz J, Schultz A, Goodrich M. Common metrics for human-robot interaction. Proc HRI ‘06. 2006:33–40.

Steriade M. Corticothalamic resonance, states of vigilance and mentation. Neuroscience. 2000;101:243–276.

Stikic Maja, Johnson Robin R, Levendowski Daniel J, Popovic Djordje P, Olmstead Richard E, Berka Chris. EEG-derived estimators of present and future cognitive performance. Frontiers in Human Neuroscience. 2011;5. doi:10.3389/fnhum.2011.00070.

Symeonidou ER, Nordin AD, Hairston WD, Ferris DP. Effects of cable sway, electrode surface area, and electrode mass on electroencephalography signal quality during motion. Sensors. 2018;18(4):1073.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93:1043–1065.

Thayer RE. Toward a psychological theory of multidimensional activation (arousal). Motivation and Emotion. 1978;2(1):1–34. https://doi.org/10.1007/BF00992729.

Thurman SM, Wasylyshyn N, Roy H, Lieberman G, Garcia JO, Asturias A, Mednick SC. Individual differences in compliance and agreement for sleep logs and wrist actigraphy: a longitudinal study of naturalistic sleep in healthy adults. PloS ONE. 2018;13(1):e0191883.

Tomaka J, Blascovich J, Kibler J, Ernst JM. Cognitive and physiological antecedents of threat and challenge appraisal. Journal of Personality and Social Psychology. 1997;73(1):63.

Tran Y, Craig A, McIsaac C.P. Extraversion–introversion and 8–13 Hz waves in frontal cortical regions. Pers Individ Differ. 2001;30(2):205–215

Unhelkar VV, Shah JA. Learning models of sequential decision-making with partial specification of agent behavior. Proceedings of the AAAI Conference on Artificial Intelligence. 2019 July;33:2522–2530.

Unsworth N, Robison MK. Pupillary correlates of lapses of sustained attention. Cogn Affect Behav Neurosci. 2016 Aug;16(4):601–615. doi: 10.3758/s13415-016-0417-4. PMID: 27038165.

Vanlalchaka RH, Zonunmawaii NPM. Correlation analysis between drowsiness and heart rate variability parameters. Advances in Engineering. 2018;178. 10.2991/msc-18.2018.8.

Van Orden KF, Jung T-P, Makeig S. Combined eye activity measures accurately estimate changes in sustained visual task performance. Bio Psychology. 2000;52:2210–400.

Van Roon, Althaus M, AM, Mulder LJ, Mulder G, Aarnoudse CC, Minderaa RB. Autonomic response patterns observed during the performance of an attention-demanding task in two groups of children with autistic-type difficulties in social adjustment. Psychophysiology. 2004;41(6):893–904.

Veltman JA, Gaillard AWK. Physiological indices of workload in a simulated flight task. Biological Psychology. 1996;42(3):323–342. https://doi.org/10.1016/0301-0511(95)05165-1.

Venek V, Scherer S, Morency LP, Rizzo A, Pestian J. Adolescent suicidal risk assessment in clinician-patient interaction. IEEE Transactions on Affective Computing. 2016.

Vicente J, Laguna P, Bartra A, Bailón R. Drowsiness detection using heart rate variability. Med Biol Eng Comput. 2016;54:927–937.

Vogels J, Demberg V, Kray J. The index of cognitive activity as a measure of cognitive processing load in dual task settings. Frontiers in Psychology, 2018.

Vossel G, Rossman R. Electrodermal habituation speed and visual monitoring performance. Psychophysiology. 1984;21:97–100.

Wang L, Jamieson GA, Hollands JG.. Trust and reliance on an automated combat identification system. Human Factors. 2009;51(3):281–291.

Wang L, Jamieson GA, Hollands JG. The effects of design features on users' trust in and reliance on a combat identification system. Proc Human Factors and Ergonomics Society Annual Meeting. 2011;55(1):375–379.

Warm JS, Finomore VS, Vidulich MA, Funke ME. Vigilance: A perceptual challenge. In: Hoffman RR, Hancock PA, Scerbo MW, Parasuraman R, Szalma JL, editors. The Cambridge handbook of applied perception research (vol. 1). New York (NY): Cambridge University Press; 2015. p. 241–283.

Warm JS, Matthews G, Finomore VS. Vigilance, workload, and stress. In: Hancock PA, Szalma JL, editors. Performance under stress. Burlington (NJ): Ashgate Publishing Company; 2008a. p. 115–141.

Warm JS, Parasuraman R, Matthews G. Vigilance requires hard mental work and is stressful. Hum Factors. 2008b;50(3):433–441.

Warm JS, Tripp LD, Matthews G, Helton WS. Cerebral hemodynamic indices of operator fatigue in vigilance. In: Matthews G, Desmond PA, Neubauer C, Hancock PA, editors. The handbook of operator fatigue. Burlington (NJ): Ashgate Publishing Company; 2012.

Watson AB, Yellott JI. A unified formula for light-adapted pupil size. Journal of Vision. 2012;12(10):12–12.

Whitham EM, Pope KJ, Fitzgibbon SP, Lewis T, Clark CR, Loveless S, Hardy A. Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. Clinical Neurophysiology. 2007;118(8):1877–1888.

Wickens CD. Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science. 2002;3(2):159–177.

Wiegmann DA, Rich A, Zhang H. Automated diagnostic aids: the effects of aid reliability on users' trust and reliance. Theoretical Issues in Ergonomics Science. 2001;2:352–367.

Wierwille WW, Wreggit SS, Kirn CL, Ellsworth LA, Fairbanks RJ. Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness. Washington (DC): Department of Transportation (US); 1994. No. HS-808 247 VPISU ISE 94-04.

Wijesuriya N, Tran Y, Craig A. The psychophysiological determinants of fatigue. International Journal of Psychophysiology. 2007;63:77–86.

Wilhelm B, Wilhelm H, Lüdtke H, Streicher P, Adler M. Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. Sleep. 1998;21(3):258–265.

Wilson Glenn F. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. The International Journal of Aviation Psychology. 2002;12(1):3–18. doi:10.1207/S15327108IJAP1201_2.

Wilson G. Applied use of cardiac and respiration measures: Practical considerations and precautions. Biological Psychology. 1992;34.

Woodworth RS, Barber B, Schlosberg H. Experimental psychology. Oxford and IBH Publishing; 1954.

Wohleber R, Matthews G, Lin J, Funke G, Neubauer C. Monitoring task fatigue in contemporary and future vehicles: a review. In International conference on applied human factors and ergonomics; 2018 July. p. 101–112.

Wohleber RW, Matthews G, Funke GJ, Lin J. Considerations in physiological metric selection for online detection of operator state: a case study. In: Schmorrow D, Fidopiastis C., editors. Foundations of augmented cognition: neuroergonomics and operational neuroscience. Springer; 2016.

Wright JL, Chen JYC, Lakhmani S.G. Agent transparency and reliability in human–robot interaction: The influence on user confidence and perceived reliability. IEEE Transactions on Human-Machine Systems. 2019;50(3):254–263.

Yanco HA, Drury J. Classifying human–robot interaction: an updated taxonomy. Proceedings of the IEEE Conference on Systems, Man and Cybernetics, The Hague, The Netherlands. 2004;3:2841–2846. New York (NY): IEEE Press.

Yang CW, Yenn TC, Lin CJ. Assessing team workload under automation based on a subjective performance measure. Safety Science. 2009;48(7):914–920.

Zhang Z, Zhang J-S. Driver fatigue detection based intelligent vehicle control. In: ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition; 2006; Hong Kong, China; IEEE Computer Society. p. 1262–1265.

Zhao X, Wei Z, Li Z, Zhang Y, Feng X. Threshold research on highway length under typical landscape patterns based on drivers' physiological performance. Discrete Dynam Nat Soc. 2015;1–15.

Zink R, Hunyadi B, Van Huffel S, De Vos M. Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. Journal of Neural Engineering. 2016;13(4):046017.

Appendix A. Documentation and Software for Human Sensing

The following link is for human sensing knowledge and capabilities within the US Army Combat Capabilities Development Command Army Research Laboratory. This is intended to serve as a resource for general knowledge and usage documentation for new and experienced users for various human sensing modalities, systems, and algorithms, to aid in multimodal physiological, behavioral, and sociological research.

This document is currently organized by sensing MODALITIES, with subsequent links to specific sensor SYSTEMS AND ALGORITHMS.

Link: https://gitlab.sitcore.net/arl/hred/human-sensing/wikis/home

Appendix B. Psychophysiological Relationship Summary

Table B-1 Notional summary of correlational and functional relationships

	Arousal	Stress	Fatigue	Workload	Vigilance	Trust
EDA tonic	+	+	*	+	*	*
EDA phasic						
HR	+	+	_	+	-	-
HRV	-	_	+	-	+	+
EEG delta power	-	*	+	+	-	*
EEG theta power tonic	-	*	+	+*	-	*
EEG theta power phasic	+	*	+*	*	*	*
EEG alpha power tonic	*	*	-	+	+	*
EEG alpha power phasic	-	*	-	*	*	*
EEG beta power	+	*	-	+*	*	*
EEG gamma power	*	*	*	*	*	*
Pupil diameter	+	*	_*	+	+	+*
PERCLOS	-	*	+	*	-	*

Note: * indicates that a relationship is largely unknown or warrants further investigation; + represents a positive correlation; - represents a negative correlation.

List of Symbols, Abbreviations, and Acronyms

ANS	autonomic nervous system
AU	Action Unit
BPM	beats per minute
CCDC	US Army Combat Capabilities Development Command
CO	cardiac output
CSF	cerebral spinal fluid
ECG	electrocardiogram
EDA	electrodermal activity
EEG	electroencephalography
EMI	electromagnetic interference
FACS	Facial Action Coding System
fEMG	facial electromyographic activity
GSR	galvanic skin response
HF	high frequency
HR	heart rate
HRV	heart-rate variability
IBI	interbeat interval
LF	low frequency
ln	natural logarithm
mRR	mean RR interval
NN	normal to normal
NN50	NN intervals greater than 50 ms
NS-SCR	non-specific SCR
nu	normalized units
PERCLOS	percentage of eye closure time
pNN50	proportion of NN50 divided by the total number of NN intervals

PSD	power spectral density
PUI	pupillary unrest index
RMSSD	root mean square of successive differences
RSA	respiratory sinus arrhythmia
SCL	skin conductance level
SCR	skin conductance response
SD	standard deviation
SDNN	SD of the NN beat intervals
SNR	signal-to-noise ratio
ULF	ultra-low frequencies
VLF	very low frequencies

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 CCDC ARL
(PDF) FCDD RLD DCI
TECH LIB

1 CCDC ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 CCDC ARL
(PDF) FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING
GROUND MD
21005-5201

1 USAF 711 HPW
(PDF) 711 HPW/RH K GEISS
2698 G ST BLDG 190
WRIGHT PATTERSON AFB OH
45433-7604

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

21 CCDC ARL
(PDF) FCDD RLH
J LANE
Y CHEN
P FRANASZCZUK
A MARATHE
K MCDOWELL
K OIE
FCDD RLH F
J GASTON (A)
FCDD RLH FA
A DECOSTANZA
C NEUBAUER
J BRADFORD
B FILES
D SPANGLER
FCDD RLH FB
D BOOTHE (A)
A OIKNINE
S THURMAN
FCDD RLH FC
K COX (A)
S GORDON
FCDD RLH FD
A FOOTS (A)
FCDD RLH FE
D HEADLEY
FCDD RLH
K SCHAEFER
FCDD RLH H
G GREMILLION